CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 111, No. 5, Pt. 1 May 2002

LETTERS TO THE EDITOR

GENERAL LINEAR ACOUSTICS [20]

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 111, NO. 5, PT. 1, MAY 2002

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 111, NO. 5, PT. 1, MAY 2002

CONTENTS—*Continued from preceding page*

ACOUSTICAL NEWS—USA

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

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

The 142nd Meeting of the Acoustical Society of America held in Ft. Lauderdale, Florida

The 142nd meeting of the Acoustical Society of America was held 3–7 December 2001 at the Ft. Lauderdale–Broward County Convention Center in Ft. Lauderdale, FL. Some events were also held at the Ft. Lauderdale Marina Marriott Hotel. This was the first time the Society has met in this city.

The meeting drew a total of 914 registrants, including 123 nonmembers and 101 students. Attesting to the international ties of our organization, 94 of the registrants (that is, about 10%) were from outside North America. There were 21 registrants from the United Kingdom, 17 from Japan, 15 from France, 8 from Germany, 4 from Brazil, 3 each from Korea and Spain, 2 each from Denmark, New Zealand, Poland, Peru, Sweden, and Tobago, and 1 each from China, Finland, India, Israel, The Netherlands, Norway, Russia, Singapore, Switzerland, Trinidad, and Venezuela. North American countries, the United States, Canada, and Mexico, accounted for 782, 34, and 4 registrants, respectively.

A total of 682 papers, organized into 84 sessions, covered the areas of interest of all 13 Technical Committees. The meeting also included 12 meetings dealing with standards. The Monday evening tutorial lecture series was continued by Tony F. W. Embleton, formerly of the National Research Council of Canada. His tutorial ''Noise Propagation and Prediction Outdoors'' was presented to an audience of about 75.

The equipment exhibit drew about ten exhibitors, and an exhibit opening reception was held on Monday evening. The number of exhibitors was fewer than usual due to current economic conditions and unfavorable timing of other conventions. Other events included the two social hours held on Tuesday and Thursday, luncheons for students in various areas of acoustics, and a Fellows Lounge which was arranged in lieu of a Fellows Luncheon. These events provided the settings for participants to meet in relaxed settings to encourage social exchange and informal discussions. The Women in Acoustics Luncheon was held on Tuesday afternoon and was attended by over 60 people.

A special fundraising dinner was sponsored by the Acoustical Society Foundation on Wednesday evening. The dinner was followed by a presentation by Sebastian Junger, author of *The Perfect Storm* and *Fire*, who recently returned from a news assignment in Afghanistan. All proceeds were added to the Foundation's endowment. Mr. Junger is the son of ASA Fellow Miguel Junger. Miguel himself was the subject of a celebration session held earlier that day and sponsored by five technical committees.

The Vern Knudsen Distinguished Lecture was presented by Anders Chr. Gade from Denmark on the topic ''Acoustic concerns related to multicultural societies.''

The Technical Committee on Signal Processing in Acoustics sponsored a series of sessions on the theory and application of time reversal and the Technical Committee on Biomedical Ultrasound/Bioresponse to Vibration organized a one-day colloquium made up of two sessions on the topic of the physics of ultrasound in relation to the biology of its therapeutic effects. These sessions were clustered early in the week, creating a form of embedded symposium on each topic. Similarly, the technical committees on Underwater Acoustics, Acoustical Oceanography, Noise, and Animal Bioacoustics sponsored a series of five sessions on acoustics research in the ocean and the welfare of marine mammals. This important program brought together ocean scientists and animal rights activists for useful and amicable exchanges. It culminated in a panel discussion chaired by the Society's Executive Director, Charles Schmid.

The plenary session included the presentation of awards and a prize. The Medwin Prize in Acoustical Oceanography was presented to Timothy G. Leighton ''for the effective use of sound in the discovery and understanding of physical and biological parameters and processes in the sea.'' Earlier in the day Dr. Leighton gave the Acoustical Oceanography Lecture, as part of the prize program (see Fig. 1).

Three science writing awards were presented. The 2000 Science Writing Award in Acoustics for a Journalist was presented to Graham Lawton for his article ''They're Playing My Song,'' which appeared in New Scientist magazine, 9 September 2000. The 2000 Science Writing Award for Professionals in Acoustics was presented to co-winners Thomas D. Rossing for his book *The Science of Musical Instruments* and to Colin Gough for his article ''The Science of the Stradivarius,'' which appeared in *Physics World*, April 2000 (see Fig. 2).

The Silver Medal in Engineering Acoustics was presented to Ilene Busch-Vishniac of Johns Hopkins University ''for development of novel electret microphones and of precision micro-electro-mechanical sensors and positioners.'' Ilene was introduced by Tony Embleton (see Fig. 3).

Election of 13 persons to Fellow grade was announced and fellowship certificates were presented. New fellows are: Shira L. Broschat, Rene Causse, Pierre Divenyi, Mathias Fink, Anthony G. Galaitis, Paul C. Hines, Anatoliy N. Ivakin, Jerry G. Lilly, Chaslav V. Pavlovic, Allan G. Piersol, Ahmet Selamet, Michael Taroudakis, and William A. Watkins (see Fig. 4).

The President expressed the Society's thanks to the Local Committee for the excellent execution of the meeting, which clearly evidenced meticulous planning. He introduced the Chair of the Meeting, Joseph M. Cuschieri

FIG. 1. ASA President William Hartmannn (r) presents the Medwin Prize in Acoustical Oceanography to Timothy G. Leighton (l).

FIG. 2. Science Writing Award winners Graham Lawton, Colin Gough and Tom Rossing (l-r).

FIG. 3. ASA President William Hartmann (l) congratulates Ilene Busch-Vishniac, recipient of the 2001 Silver Medal in Engineering Acoustics.

FIG. 4. Newly elected ASA Fellows with ASA Vice President Janet Weisenberger and President William Hartmann (c) at the Plenary Session.

FIG. 5. ASA President William Hartmann congratulates Joseph Cuschieri, General Chair of the meeting, on a well-organized meeting.

(see Fig. 5), who acknowledged the contributions of the members of his committee including: Stewart A. L. Glegg, Technical Program Chair; Robert Coulson, Audio-Visual; Pierre Beaujean, Hotel/Facilities; Natasha Walker, Signs; Janette Garcia, Accompanying Persons Program. He also extended thanks to the members of the Technical Program Organizing Committee: Stewart A. L. Glegg, Technical Program Chair; David Palmer and Michael Brown, Acoustical Oceanography; Joseph E. Blue, Edward Gerstein, and David Mann, Animal Bioacoustics; Lily Wang, Architectural Acoustics; Subha Maruvada, Biomedical Ultrasound/Bioresponse to Vibration; James Cottingham, Education in Acoustics and Musical Acoustics; Elizabeth McLaughlin, Engineering Acoustics; Mahlon D. Burkhard, Noise; Kerry Commander, Physical Acoustics; Richard M. Stern, Psychological and Physiological Acoustics; David Chambers, Signal Processing in Acoustics; Betty Tuller and Gautam Vallabha, Speech Communication; Courtney Burroughs, Structural Acoustics and Vibration; and Ellen Livingston and John Perkins, Underwater Acoustics.

WILLIAM M. HARTMANN
President 2001–2002

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.

2002

3–7 June	143rd Meeting of the Acoustical Society of America, Pittsburgh, PA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org;WWW:asa.aip.org].
19–21 Aug.	INTER-NOISE 2002, Dearborn, MI [INTER-NOISE 02 Secretariat, The Ohio State University, Department of Mechanical Engineering, 206 West 18th Ave., Columbus, OH 43210-1107; E-mail: hp@internoise2002.org].
2–6 Dec.	Joint Meeting: 144th Meeting of the Acoustical Society of America, 3rd Iberoamerican Congress on Acoustics, and 9th Mexican Congress on Acoustics, Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org/cancun.html].

2003

28 April–2 May	145th Meeting of the Acoustical Society of America, Nashville, TN [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
10–14 Nov.	146th Meeting of the Acoustical Society of America, Austin, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW:asa.aip.org].

2004

24–28 May	75th Anniversary Meeting (147th Meeting) of the Acoustical Society of America, New York, NY [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
13–20 Nov.	148th Meeting of the Acoustical Society of America, San Diego, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

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

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

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.

NHCA's 27th Annual Hearing Conservation Conference

On 21–23 February 2002 in Dallas, TX, the National Hearing Conservation Association (NHCA) held its 27th Annual Hearing Conservation Conference, chaired by ASA member Mary McDaniel, Vice President of NHCA. This year, as it was three times in the past, the Acoustical Society of America was a conference affiliate. Daniel L. Johnson, Standards Director, and Susan Blaeser, Standards Secretariat, represented ASA at the exhibits and other functions.

The conference, attended by approximately 260 hearing conservation professionals, included four concurrent half-day workshops from which attendees could select two. The workshops covered the epidemiology of noise induced hearing loss including oxidative mechanisms, and pharmacologic protective agents, workers' compensation, tinnitus, and metrics for evaluating hearing conservation programs, and an all-day beginner's seminar on the basics of hearing loss prevention. The program, diverse as usual, began with a retrospective lecture by Alice Suter on little-known things that happened along the way as OSHA (Occupational Health and Safety Administration) developed and implemented noise regulations, and on the second day included the second installment of the Don Gasaway Lecture, delivered by Ted Madison, on the topic of communicating the value of hearing. Other aspects of the broad-ranging program were an encore presentation of NHCA's practical pop-ups (short, 10-minute to-the-point presentations), nine posters, two forums, a series of round-table breakfast discussions, and a number of allied committee and ANSI working group meetings. The luncheon lecture, ''Orality and Literacy,'' was delivered by Kent Menzel, a professor in communications at DePauw University. Of course, there was ample time to visit and socialize during the exhibitor receptions and the live auction, at which a fun time was had by all.

The lectures covered topics such as a discussion by Bob Dobie on how changes in audiometric configuration help determine whether a standard threshold shift is work-related, an interesting presentation by Beth Cooper and Dick Danielson on hearing conservation for the international space station, an exciting point–counterpoint session by Elliott Berger of EAR and Dan Gauger of Bose discussing when active noise reduction is useful for hearing protection technology and when it is not, augmented by auditory demonstrations of actual hearing protector attenuation characteristics, and an important paper by Jennifer Tufts on the effect of wearing hearing protectors on the production of speech in noise. The posters were equally as varied, covering issues like probe microphone measurements of hearing protector performance, an unusual paper entitled ''It if sounds 'delicious' it might be too loud,'' and the fun, motivational efforts of Deanna Meinke, James Lankford, and Laurie Wells related to the collection of favorite sounds.

Long-time NHCA contributor and Past President, Merlyn Lubiens, received the gratitude of the organization when he was presented the Michael Beall Threadgill Award for Outstanding Leadership and Service to NHCA.

Loose-leaf proceedings are available from the NHCA Executive Offices in Denver, CO (303-224-9022, email nhca@gwami.com), and abstracts can be viewed on the NHCA website at www.hearingconservation.org. The next NHCA Annual Hearing Conference will be held in Dallas on 20–22 February 2003.

ELLIOTT H. BERGER

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).]*

Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation

Dafydd Gibbon, Inge Mertins, and Roger Moore, Editors

Kluwer Academic Publishing, The Netherlands. 2000.
Price: $175.00 (hardcover) ISBN: 0792379047.

This is not a traditional book review—because this is not a traditional technical text. It is titled a Handbook, and in every sense it is that. One seldom reads such from front cover to back, and neither have I. Rather I have perused the volume, concentrating mainly on topics of personal interest, to savor the rich mixture of views and data provided.

The Editorial Preface well sets the context for the document. It addresses Language Technologies—with the aim of providing best practices for evaluating products, for applying language terminology, for conducting research on multimodal and audiovisual systems, and for utilizing database resources. It constitutes a large compendium, based primarily on European research. For this reason it will be particularly valuable to researchers who have limited exposure to work in Europe.

The document is the product of an activity funded by the European Union to promote the use of standards in spoken language processing, text, and terminology—conducted under the aegis of the multinational Expert Advisory Group on Language Engineering Standards (EAGLES). It lists some 20 scientists as ''Main Technical Authors,'' along with acknowledgments to a comprehensive list of collaborators across European labs, universities, institutes, and companies. The editors, Dafydd Gibbon (U. Bielefeld), Inge Mertins (U. Bielefeld), and Roger Moore (DERA and 20/20 Speech Limited), are also contributors and key participants in EAGLES projects. The book is dedicated to Dr. Christian Benoît, one of the lead technical contributors and a cherished colleague, who died before the project completed.

Although ''Multimodal'' appears first in the title, most of the book focuses on spoken language. As conversational interfaces have advanced, automatic speech recognizers and large speech databases (with which to train statistical classifiers favored in current designs) have almost become commodities. Speech understanding and dialog design have not. The latter represent major research frontiers. Even speech recognizers continue to present substantial challenges on two fronts: (*i*) in elevating recognition accuracy for spontaneous speech (typical of interactive exchanges, which often are nongrammatical); and (*ii*) in reducing the onerous (and expensive) training of classifiers by exposing them to vast amounts of target speech. In both areas of research—understanding and recognition—this Handbook has much to contribute. And in both these areas, a continuing focus is to build solutions rapidly and economically as the application tasks and the languages change.

As a personal bias, I would have liked more in the book on Multimodal techniques. This area has not been researched nearly as intensely as speech recognition. It is becoming increasingly important, as the push becomes stronger to transcend mouse and keyboard interfaces and achieve more natural communication with machines (particularly employing the simultaneous sensory modes of sight, sound, and touch). In addition to the need for new interface devices and transducers for sensory modes, a clear challenge that should attract the proponents of this book is the need for a ''language framework'' for multimodal communication. This implies software agents that can fuse simultaneous multisensory inputs to machines (for example, ''Move this to there''—signaled by speech, manual gesture, and possibly visual confirmation). A multimodal language framework, similar to that developed over many years for spoken language, would enable multimodal parsing and application of semantic analysis and contextual constraints—all aimed at providing the machine a reliable estimate of user intent, along with input to dialog generation modules that produce a text-to-synthetic voice response.

A good part of the book familiarizes the reader with database resources that may be widely helpful—as well as with best practices for their use. And, it has a glossary appended for the many, many abbreviations that are used throughout. This does not always suffice though (even in a book devoted to terminology), and the reader should be prepared for some esoteric (if not intimidating) verbiage. Exposition of the EAGLET Term Databank provides a sample (p. 271): ''In order to take the Scylla of heterogeneity in the field of SL terminology into account, and avoid, on the other hand, the Charybdis of completely *ad hoc* hybrid description, a new approach is proposed which combines the traditional semasiological and onomasiological approaches to terminology characterization,—(etc).''

Discharging the implicit obligation of a reviewer to nit-pick, other minor annoyances can be cited. The physical realization might be better. Most of the text is relatively small font (an estimate is 9 point Times New Roman) and lightly, or delicately, printed. Section headings are not particularly emphasized, and do not stand out as they might if bold font were used. Further on the side of verbosity, discussions that belabor some controversial views—such as distinguishing the terms ''multimodal'' and ''multimedia'' (p. 105)—become a bit wearisome. But none of these things detract from the scientific value—they just place a small burden on the reader (and might even stimulate resolve in the serious student).

Again, the Handbook is perhaps not a traditional one (say, where one looks up the speed of sound in esoteric media), but it provides definitions, terminology, delineations of research, and extensive pointers to significant work, especially that of the European community. Its attention to ''product and service evaluation'' appropriately focuses on techniques and procedures, not on specific products or manufacturers. The main body of text runs to 327 pp. The extensive Bibliographical References and Appendices fill the remainder of the 519 pp. total. It is accompanied by a CD-ROM version with a disclaimer relative to support and warranty. The book is offered as a companion to the earlier *Handbook of Standards and Resources for Spoken Language Systems* (Gibbon *et al.*, 1997), produced also under the EU aegis. This reviewer does not know the companion, but is herewith prompted to seek it.

In sum, the European community, and especially the editors and contributors for this comprehensive volume, have provided a valuable service to the scientific academy in this compilation. I doubt that I would ask each of my graduate students to obtain a copy, but I am certainly glad we have one in the lab. The book seems to be an excellent exemplar of the burgeoning spirit of cooperation across Europe—a spirit not limited solely to science, and a model that may eventually pervade the global community. Finally, those of us who are very poor linguists appreciate the choice of English for the publication.

JIM FLANAGAN
Center for Advanced Information Processing
Rutgers University
Piscataway, New Jersey 08854-8088

 0001-4966/2002/111(5)/1962/1/$19.00

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
IBRAHIM M. HALLAJ, *Wolf Greenfield & Sacks P.C., 600 Atlantic Avenue, Boston, Massachusetts 02210*
DAVID PREVES, *4 Deerfield Drive, Princeton Junction, New Jersey 08550*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*

6,305,233

43.20.Mv DIGITAL SPEED DETERMINATION IN ULTRASONIC FLOW MEASUREMENTS

Colin Walter Braathen *et al.*, assignors to Commonwealth Scientific and Industrial Research Organisation; AGL Consultancy Pty Limited
23 October 2001 (Class 73/861.28); filed in Australia 19 October 1995

A method is disclosed for determining flow velocity in a tube using a transmit/receive pair. A reference pulse packet is received by a receiver and digitized. Several segments of the digitized signal are then analyzed on their slopes in order to identify and match the received signal with a copy of the transmitted signal. The matching allows determination of signal travel time and other information upon comparison of selected features in the sent/received signals.—IMH

6,311,573

43.20.Rz ULTRASONIC TRANSDUCER FOR HIGH TRANSDUCTION IN GASES AND METHOD FOR NON-CONTACT ULTRASOUND TRANSMISSION INTO SOLID MATERIALS

Mahesh C. Bhardwaj, Boalsburg, Pennsylvania
6 November 2001 (Class 73/866.5); filed 19 June 1997

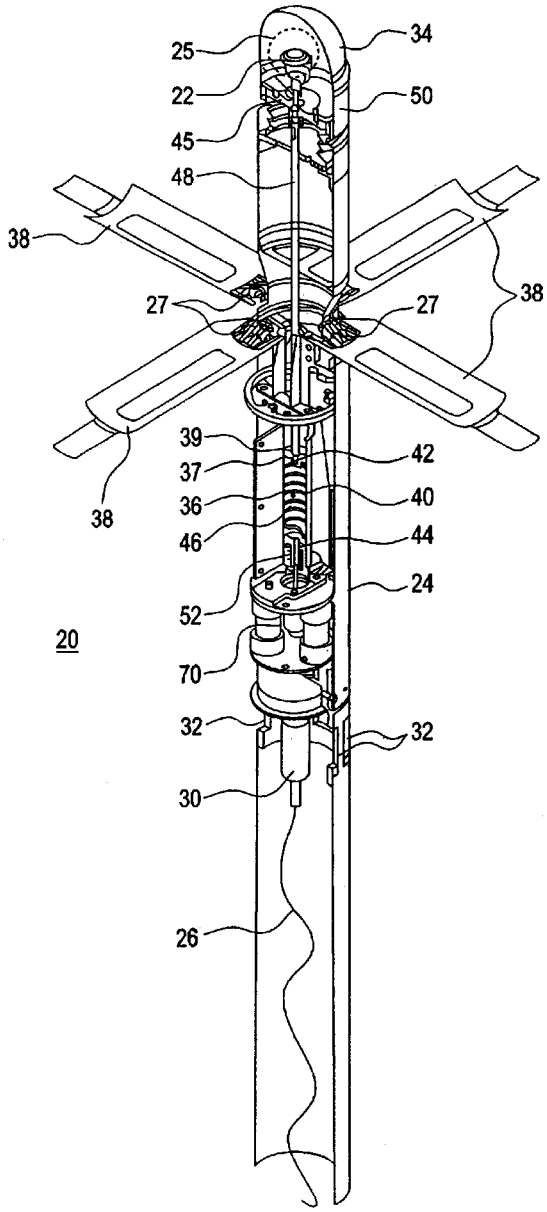
This efficient transducer system for use with gaseous media uses an active transducer having a polymer facing layer to which a final fibrous layer is bonded. The fibrous layer may be made of wood, cotton, fabric, felt, etc. Care is given not to allow the bonding agent to penetrate too deeply into the fibrous layer to avoid destroying its open fiber qualities.—IMH

6,307,810

43.28.Tc UNDERWATER LAUNCHED ACOUSTIC WARNING ASSEMBLY

Ofir Shany *et al.*, assignors to Rafael-Armanent Development Authority Limited
23 October 2001 (Class 367/131); filed in Israel 31 August 1998

A floatable transceiver assembly is described, launchable from an underwater vehicle, used to monitor air-borne sounds, and, in turn, to relay them to the underwater vehicle. With reference to the figure, which is a



partial cross-sectional view, the cylindrical housing **24** supports at its upper end a microphone **22**. Element **25** is a windscreen while cover **34** is ejected after the upper portion of the unit has surfaced. Fins **38** retard the amount the unit will broach the surface after launch. Communciation with the launch vehicle can be via wire link **26** which is payed out from spool **30**, although an acoustic link is also suggested. In addition to power supply and necessary electronics, the unit also houses a self-destruction mechanism. The underwater vehicle contains the necessary equipment for receiving the signals from the transceiver assembly and for identifying the source of the original airborne sound, e.g., a helicopter, by matching with prerecorded signatures.—WT

6,304,514

43.30.Gv ULTRASONIC MEASURING DEVICE WITH TRANSMITTERS AND RECEIVERS FOR LOCATING THE GEOMETRIC POSITION OF THE BORDER BETWEEN A FIRST AND SECOND MATERIAL FROM A REFERENCE LOCATION

Torsten Schulze, assignor to Theyson GmbH
16 October 2001 (Class 367/98); filed in Germany 20 September 1997

Of interest to those using transmit-receive/pulse-echo techniques to determine the thickness of a material or the location of a defect, this patent is directed to a technique for receiving an echo from a far interface of a material and subtracting this signal from the corresponding echo received from the near interface of the material. The time difference between rise times is taken, identifying the travel time of the pulse through the material. The patent might carry the influence of a German parent, as it contains just one claim (drawn to the apparatus) and the Detailed Description is all of two paragraphs long.—IMH

6,304,513

43.30.Wi METHOD FOR CORRECTING EFFECTS OF AN ANTENNA INTERFERING MOVEMENTS IN A SONAR WITH SYNTHETIC ANTENNA

Didier Billon, assignor to Thomson Marconi Sonar S.A.S.
16 October 2001 (Class 367/88); filed in France 7 October 1997

A method is discussed for obviating the negative effects upon the localization ability of a synthetic sonar array caused by spurious motions of the axially oriented, submarine-mounted, sonar array. These corrections are achieved by cross correlating successive signals and by using information about the roll of the array obtained from a rate gyro and about the angle of elevation of the received signal determined from a small auxiliary array mounted perpendicularly to the primary array.—WT

6,310,830

43.30.Wi ENVIRONMENTALLY ADAPTIVE SONAR SYSTEM

Henry M. Gruen, assignor to Northrop Grumman Corporation
30 October 2001 (Class 367/88); filed 7 April 2000

Improved detection and classification of targets in shallow waters is said to be realized by adaptively modifying the vertical beamwidth of the receive beam of a sonar system as a function of water depth, altitude of the sonar, and range to the target. This modification is accomplished by dynamically controlling the number of active elements in the hydrophone array.—WT

6,302,857

43.35.Wa METHOD AND APPARATUS FOR LOWERING AUDIBLE NOISE EMISSIONS FROM LITHOTRIPTORS

Vince Landeck, St. Charles, Missouri
16 October 2001 (Class 601/4); filed 2 June 2000

Lithotriptors are external medical and therapeutic devices used to disintegrate concretions (e.g., kidney stones) induce bone growth, and treat soft tissues. These devices operate by producing focused acoustic shock waves, capable of generating extremely high-pressure differentials at localized regions within a patient's body, that act upon targeted concretions or body tissue being treated. Noise is naturally generated by the generation of shock waves, and the patent describes a noise attenuator in the form of a noise-damping shroud that is configured to surround the generator of a lithotriptor. The shroud consists of an insulating body and a cover. The insulating body consists of sound-damping insulation that is configured to enclose all of the housing that typically encases the generator of the lithotriptor. The cover is generally a thin pliant material surrounding the insulating body.—DRR

6,312,402

43.35.Wa ULTRASOUND CATHETER FOR IMPROVING BLOOD FLOW TO THE HEART

Douglas R. Hansmann, assignor to Ekos Corporation
6 November 2001 (Class 604/22); filed 24 September 1998

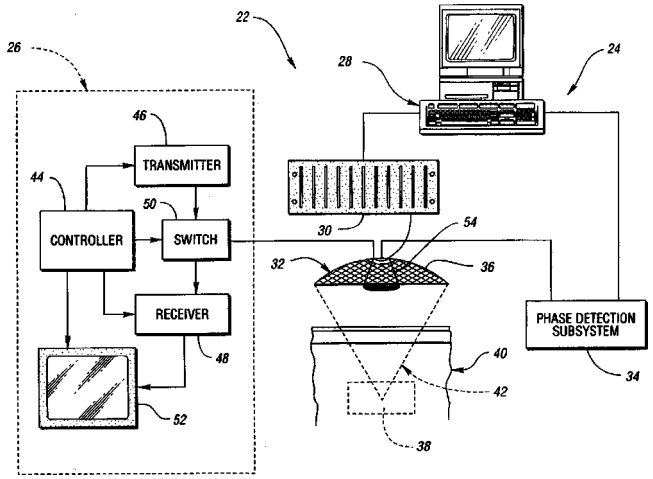
The purpose of this ultrasonic catheter is multifold: it is intended to improve the blood flow to a patient's heart, to create revascularization channels in a patient's heart, and to provide an energy source which in combination with a medicament stimulates angiogenesis in the heart. The catheter has an elongated catheter body with a proximal end and a distal end, thus constituting a catheter lumen. The elongated catheter body incorporates an introducer distal portion with a tissue-piecing distal end. An ultrasound transducer is positioned in the introducer distal portion.—DRR

6,309,355

43.35.Wa METHOD AND ASSEMBLY FOR PERFORMING ULTRASOUND SURGERY USING CAVITATION

Charles A. Cain and J. Brian Fowlkes, assignors to The Regents of the University of Michigan
30 October 2001 (Class 600/439); filed 22 December 1998

This is a method for using an ultrasonic beam to generate a controlled surgical lesion in a selected treatment volume within a patient. First, multiple microbubbles are created in the treatment volume. Preferably, the threshold for cavitation of microbubbles should be below the cavitation



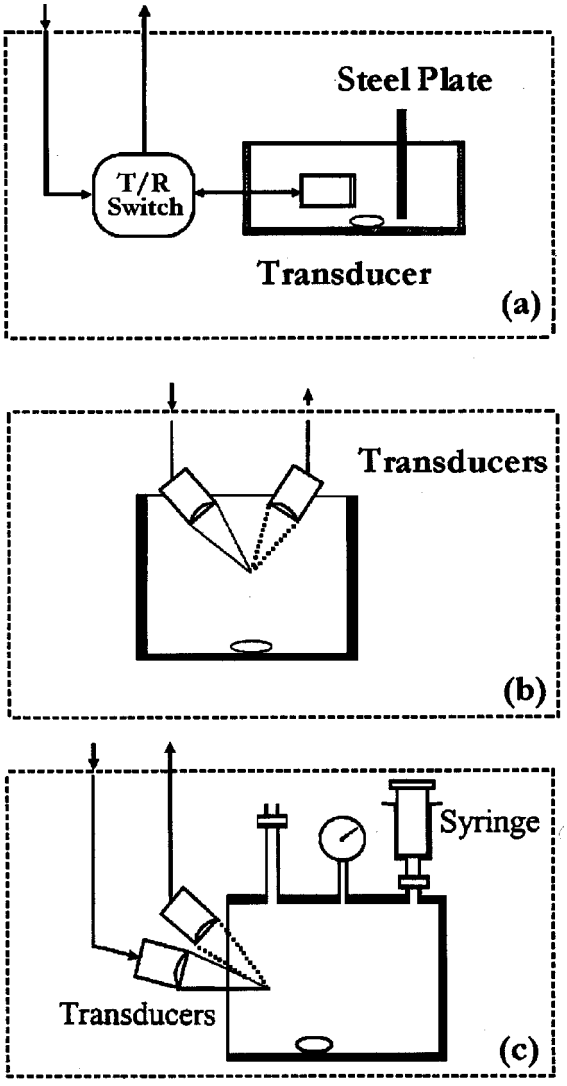
threshold for the surrounding tissues. The expected location of the surgical-lesion may be previewed before the microbubbles are cavitated with the ultrasound to create the controlled surgical lesion. The creation of the surgical lesion can then be verified. It is claimed that the cavitation threshold is predictable and that a low-frequency ultrasound beam may be used to cavitate the microbubbles without causing damage to the surrounding tissues.—DRR

6,302,845

43.35.Yb METHOD AND SYSTEM FOR PRESSURE ESTIMATION USING SUBHARMONIC SIGNALS FROM MICROBUBBLE-BASED ULTRASOUND CONTRAST AGENTS

William Tao Shi *et al.*, assignors to Thomas Jefferson University
16 October 2001 (Class 600/438); filed 20 March 1998

The object of the methodology described here is to provide a direct noninvasive measurement of hydrostatic pressure inside the heart, in other organs, and in major blood vessels. The system uses ultrasound contrast-agents containing microbubbles to achieve noninvasive subharmonic-aidedpressure estimation in the cavities of the heart, organs, and principal blood vessels. Some contrast agents are particularly well suited for pressure measurements because their substantial compressibility enables the microbubbles to vary appreciably in size in response to changes in pressure. Pressure changes, in turn, affect the reflectivity of microbubbles after intravenous injection of a contrast agent. Such microbubbles can, e.g., increase the backscatter from blood. The nonlinear properties of these microbubbles may be utilized to create new harmonic and subharmonic modalities.—DRR

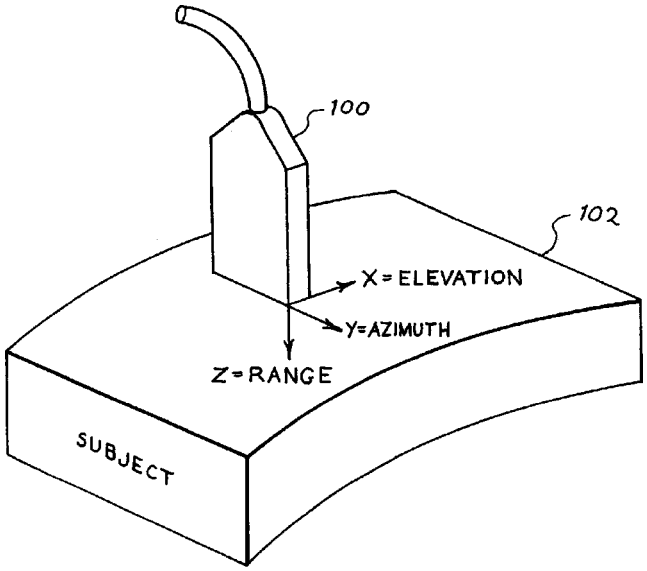


6,306,091

43.35.Yb DIAGNOSTIC MEDICAL ULTRASOUND SYSTEMS AND METHODS UTILIZING ESTIMATION OF 3-DIMENSIONAL RIGID BODY TRANSFORMATION

Thilaka S. Sumanaweera and John A. Hossack, assignors to Acuson Corporation
23 October 2001 (Class 600/443); filed 6 August 1999

The patent refers to medical diagnostic systems and methods capable of constructing three-dimensional images from disjoint two- or three-dimensional image data sets of humans or animals. More specifically, the method entails ultrasound systems that provide accurate three-dimensional reconstruction between any two ultrasonic data sets acquired by translating



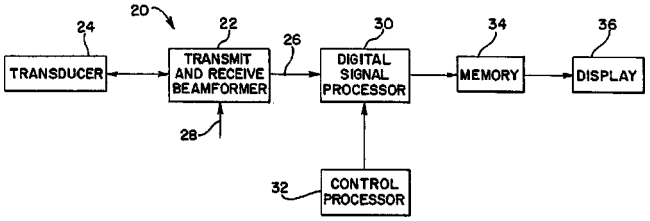
or rotating a transducer along one or more of the six degrees of freedom. In various embodiments, either quaternions or orthonormal matrices are used to execute the affine transformation. It is claimed that the resultant three-dimensional reconstructions are rendered more quickly with greater accuracy and with less deformation or skewedness.—DRR

6,306,093

43.35.Yb METHOD AND SYSTEM FOR ULTRASOUND ENHANCED-RESOLUTION SPECTRAL DOPPLER

John S. Wang, assignor to Acuson Corporation
23 October 2001 (Class 600/454); filed 22 May 1997

This method and system is claimed to provide an efficient method for processing and displaying various ultrasound data. In one embodiment, a method for generating data for a spectral Doppler strip in an ultrasound system is provided. A beam-former provides Doppler data to a signal



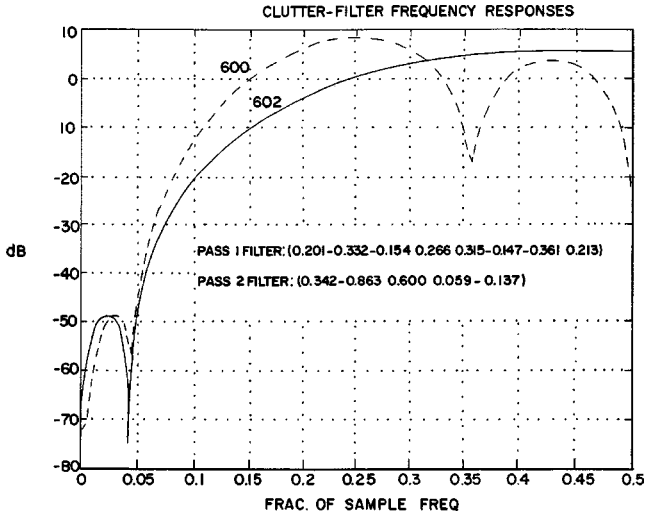
processor. The signal processor applies an autoregressive model and determines autoregressive parameters responsive to the Doppler data. Information responsive to the autoregressive parameters is generated; and a window function is applied to the information.—DRR

6,309,357

43.35.Yb MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD FOR IMPROVED FLOW OR MOVEMENT DETECTION WITH MULTIPLE CLUTTER FILTERS

Ismayil M. Guracar and Patrick J. Phillips, assignors to Acuson Corporation
30 October 2001 (Class 600/454); filed 9 August 1999

Ultrasound systems image blood flow and tissue movement using correlation or Doppler techniques. The flow or movement is represented by one or more of various estimated parameters, such as energy, velocity, and/or variance. Prior to the estimation of these parameters, a clutter filter may be used to suppress undesired signals, such as those associated with reflections



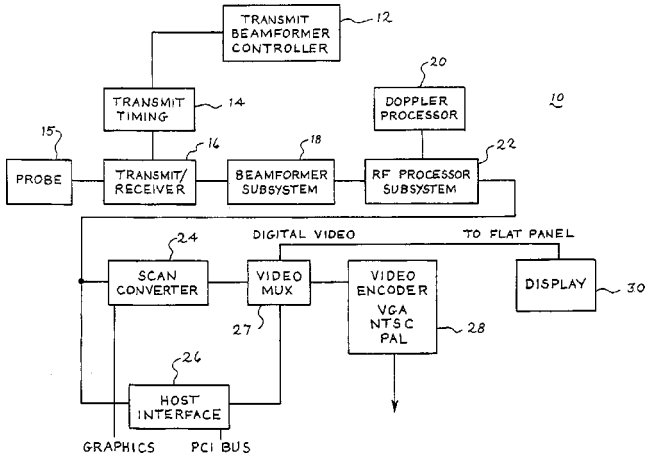
of ultrasonic energy from stationary or slow-moving tissue. In the system described in this patent, more than one clutter filter is used. Each clutter filter's frequency response is optimized differently. Estimates of the flow or movement are generated from the output of the clutter filters. Using selection or combination of the resulting estimates, the best attributes of each filter are used to produce images.—DRR

6,312,381

43.35.Yb MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD

Christopher B. Knell *et al.*, assignors to Acuson Corporation
6 November 2001 (Class 600/437); filed 14 September 1999

This patent, saddled with a most unspecific title, deals with a system for providing temporally accurate EKG data linked to ultrasonic imaging



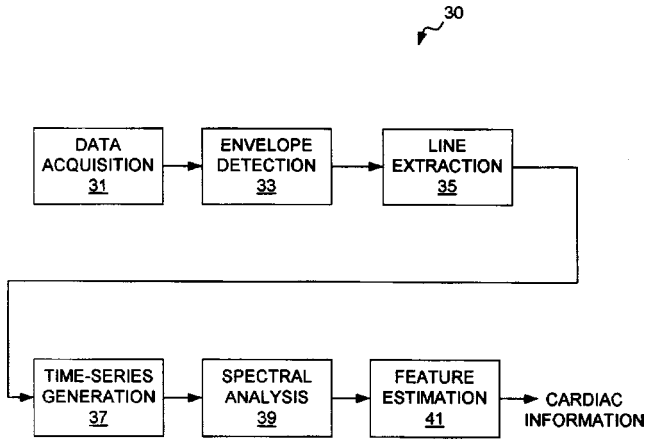
data. The frame of imaging data contains EKG data in a line of information. Additional EKG data may be provided separately from the imaging data. A scan converter processes both the EKG data and imaging data at virtually the same time, maintaining temporal accuracy. The EKG data provided constitutes an entire trace so that the trace on the display appears to be smoothly updated.—DRR

6,312,382

43.35.Yb METHOD AND APPARATUS FOR EXTRACTING CARDIAC INFORMATION FROM ACOUSTIC INFORMATION ACQUIRED WITH AN ULTRASOUND DEVICE

Ronald Mucci, Westwood, Massachusetts *et al.*
6 November 2001 (Class 600/437); filed 15 November 1999

The patent pertains to ultrasound imaging, and, more particularly, to an apparatus for processing acoustic information from an ultrasonic device for monitoring cardiac information, such as a patient's heart rate. An ultrasonic system is applied to image a patient's heart in order to obtain cardiac information. The acoustic information is converted into electrical signals. These



are used to generate a time series of image frames which are analyzed by Fourier analysis to determine the fundamental frequency corresponding to the patient's heart rate. While it seems that this is a rather elaborate way of measuring the heart rate, which can be done by feeling a pulse or utilizing a

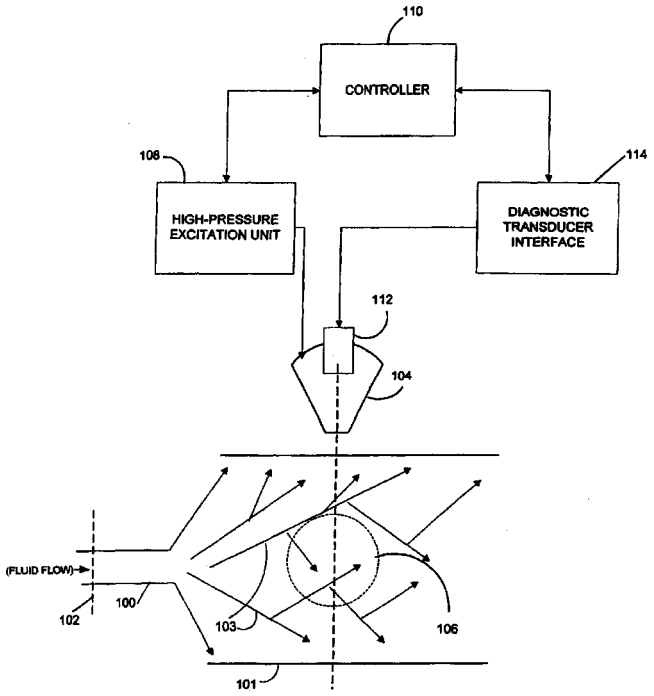
stethoscope, it may be possible that other embodiments of this apparatus can be used to extract additional information on the cardiac status.—DRR

6,312,383

43.35.Yb DUAL BAND ULTRASONIC SYSTEMS

Frederic Louis Lizzi and Cheri Xiaoyu Deng, assignors to Riverside Research Institute
6 November 2001 (Class 600/437); filed 26 May 1998

This patent describes a hand-held device that measures blood flow rate and perfusion through the use of an ultrasound contrast. The device includes a source that generates a focused pulse of ultrasound energy at a frequency and magnitude sufficient to modify the contrast agent particles in a target area in the tissue region. A second ultrasound signal is at a



frequency and magnitude which do not substantially modify the contrast agent. A controller operates the source to determine an initial measurement of the contrast agent in the target area, to modify the contrast agent, and to monitor the level of the contrast agent. The processor calculates the time required to restore the contrast agent level to the initial level and establishes a perfusion rate therefrom.—DRR

6,298,141

43.38.Ja METHOD AND APPARATUS FOR AUDIO BASS ENHANCEMENT IN AN ELECTRONIC DEVICE

Scott N. Hickman, assignor to Hewlett-Packard Company
2 October 2001 (Class 381/333); filed 30 October 1997

Using vibrations in the housing of a portable computer to enhance audio bass response is not a new idea. The interesting feature of this patent is its description of a miniature shaker motor used to generate the low-frequency vibrations.—GLA

6,298,140

43.38.Ja ELECTROACOUSTIC TRANSDUCER WITH IMPROVED TONAL QUALITY

Christos Manavopoulos, Refina Attika, Greece
2 October 2001 (Class 381/152); filed 20 February 1998

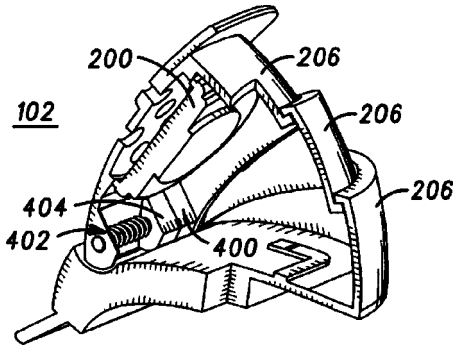
This patent pertains to flat-panel loudspeakers composed of arrangements of active elements suspended in baffles. Openings in the baffles permit some of the radiation from the backs of some elements to cancel partially the radiation from the fronts of the elements, smoothing the frequency response of the system. Asymmetry and nonuniform driving of the active elements increase the number of excited modes, further smoothing the radiated spectrum.—EEU

6,292,573

43.38.Ja PORTABLE COMMUNICATION DEVICE WITH COLLAPSIBLE SPEAKER ENCLOSURE

Robert A. Zurek *et al.*, assignors to Motorola, Incorporated
18 September 2001 (Class 381/386); filed 30 September 1999

A number of earlier patents describe nesting or telescoping loudspeaker enclosures. This invention is a variant intended specifically for use in very small, portable communication systems. Several possible methods of achieving a good air seal are suggested, but it is more than likely



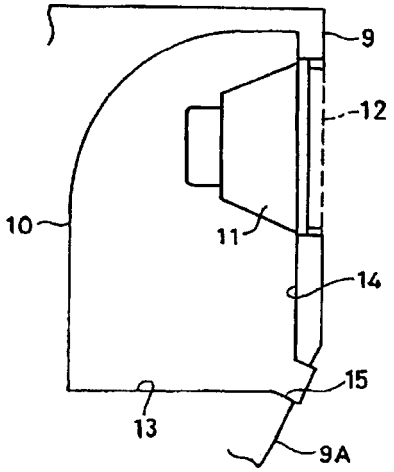
that substantial air leaks will develop after repeated opening and closing. A practical alternative might be to incorporate controlled resistive leakage calculated to smooth the response of a typical small, underdamped speaker.—GLA

6,298,943

43.38.Ja BASS-REFLEX SPEAKER ASSEMBLY

Katsuhisa Yamada *et al.*, assignors to Honda Giken Kogyo Kabushiki Kaisha
9 October 2001 (Class 181/156); filed in Japan 1 December 1998

Large cruising motorcycles, such as the Honda Gold Wing, often include elaborate stereo systems. In this application, loudspeakers are exposed to rain and fluctuating air pressure. By mounting each loudspeaker in a



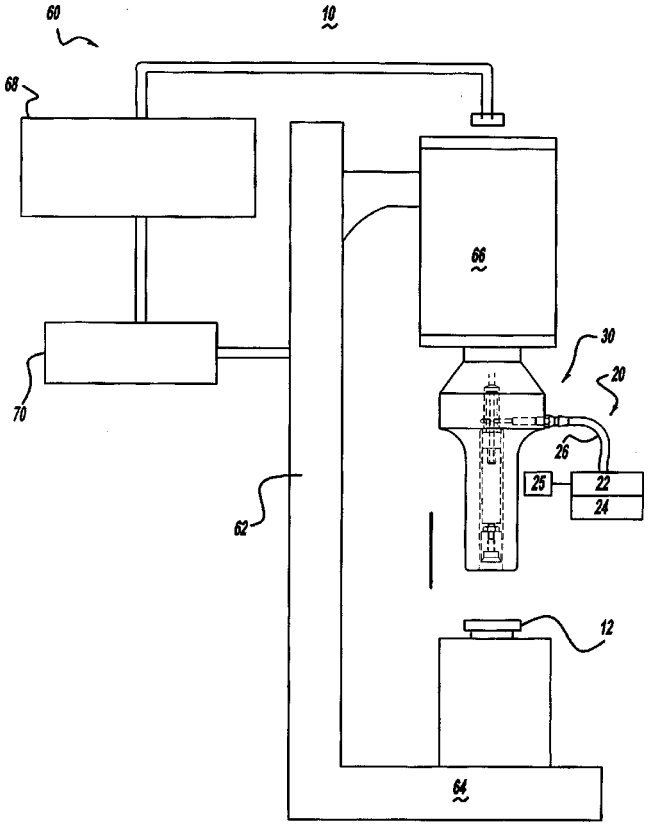
vented enclosure, front and rear air pressures are equalized. Moreover, if vent duct **15** is located as shown, it can also serve as a drain for any water that finds its way inside.—GLA

6,309,490

43.38.Ja AIR ACTUATED ULTRASONIC TOOL

Shawn K. Davis and John Ablamsky, assignors to Branson Ultrasonics Corporation
30 October 2001 (Class 156/73.3); filed 10 May 2000

An ultrasonic horn, e.g., for welding plastic parts, is combined with an



air-pressure controllable-force actuator to control the application of ultrasound from the horn to the parts.—IMH

6,301,034

43.38.Kb PULSED LASER MICROPHONE

John R. Speciale, Wonder Lake, Illinois
9 October 2001 (Class 359/151); filed 22 October 1997

Pulsed laser light is reflected from a moving diaphragm and focused onto an optical receiver. According to the patent text, the modulated, pulsed output can be converted directly to a digital audio signal by a simple comparator circuit. If desired, transmission and reception can both be conveyed via fiber optic cable so that the microphone itself contains no active elements.—GLA

6,297,695

43.38.Lc HIGH VOLUME EXPANDER CIRCUIT

Wayne M. Schott, assignor to U.S. Philips Corporation
2 October 2001 (Class 330/110); filed 16 September 1999

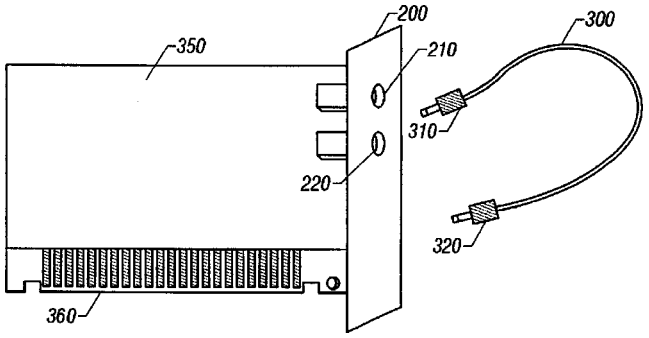
This might better be described as an unlimiter. The idea is to partially undo the heavy audio limiting imposed on typical signals from AM radio, VCR sound tracks, and the like. The gain of an opamp remains fixed until the input signal exceeds a predetermined level, at which point it increases to a second gain setting. The patent does not address the question of matching the trigger level to a particular program source.—GLA

6,304,865

43.38.Md AUDIO DIAGNOSTIC SYSTEM AND METHOD USING FREQUENCY SPECTRUM AND NEURAL NETWORK

Alan K. Christensen and Christopher F. Broadbent, assignors to Dell U.S.A., L.P.
16 October 2001 (Class 706/20); filed 27 October 1998

This patent describes a test setup for evaluating and calibrating computer sound cards. A test signal is recorded by the sound card and then played back into the test system. A Fourier transform of the reproduced



signal is processed by a neural network trained to recognize any discrepancy from the expected signal analysis.—DLR

6,296,926

43.38.Ne EMBOSSED SOUND TRACK

Stefan Huebner, Munich, Germany
2 October 2001 (Class 428/172); filed in Germany 18 August 1997

Talking greeting cards and cereal boxes rely on an embossed strip as the recording medium—a kind of three-dimensional bar code. The patent describes an improved recording material and method which "...permits high reproduction quality, has sufficient life, is easy to handle, and can be produced efficiently."—GLA

6,295,364

43.38.Si SIMPLIFIED COMMUNICATION SYSTEM

Brian M. Finn *et al.*, assignors to Digisonix, LLC
25 September 2001 (Class 381/110); filed 30 March 1998

A live, two-way intercom connection between two acoustic spaces is subject to feedback in the audio loop. This patent does not address that issue, normally solved by limiting the live path to one side at a time, in what is known as a simplex system. However, if the spaces are acoustically coupled, feedback may still occur. For intended application between the front and back seat of a car, this system introduces high-pass and equalization filters in both microphone pathways. For the front-to-back path, the high-pass filter is dynamically adjusted by a control signal from the vehicle speedometer.—DLR

6,292,570

43.38.Vk SURROUND SOUND

Ronaldus M. Aarts, assignor to U.S. Philips Corporation
18 September 2001 (Class 381/18); filed in the European Patent Office 13 February 1998

This patent addresses the problem of reproducing surround sound through TV speakers or computer sound systems in which surround speakers are absent. Much of the previous work in this field makes use of interaural cancellation and/or pinnae filter simulation. The required circuitry is complicated and the effective listening area is constricted. The inventors propose a much simpler approach in which the surround channel is divided into two frequency bands, ''expanded,'' and then recombined with the main left and right channels.—GLA

6,301,367

43.38.Si WEARABLE AUDIO SYSTEM WITH ACOUSTIC MODULES

James H. Boyden *et al.*, assignors to Interval Research Corporation
9 October 2001 (Class 381/376); filed 8 March 1995

This patent describes wearable earphones; that is, small loudspeakers held in proximity to the user's ears. Each module has an outlet port adjacent to the ear and a vent port located far enough away to minimize low frequency cancellation. The inventors must have faith in their design because the final patent is a continuation in part of a continuation of a division of an application going back to 1995.—GLA

6,289,735

43.40.Le MACHINE DIAGNOSTIC SYSTEM AND METHOD FOR VIBRATION ANALYSIS

Carl J. Dister *et al.*, assignors to Reliance Electric Technologies, LLC
18 September 2001 (Class 73/579); filed 29 September 1998

A vibration sensor is attached to a dynamoelectric machine which is to be monitored. The sensor signal is analyzed to obtain a frequency domain vibration signature. The processor logic then scans this signature at several harmonics of a known frequency associated with a machine component of concern, such as a frequency associated with the balls in a bearing, and compares the amplitudes of adjacent harmonics. The shape and magnitude of the largest of a set of harmonics, thought to correspond to a resonance in the transmission path between the machine component and the sensor, are analyzed to evaluate the condition of the machine.—EEU

6,301,490

43.38.Si AUDIO HEADSET COMMUNICATION APPARATUS AND METHOD

Edward William Callan, San Diego, California
9 October 2001 (Class 455/568); filed 12 June 1998

Suppose that you are a switchboard operator or a dispatcher. Your job requires simultaneous voice communication with two or more remote stations. If you wear a stereo headset, then person A can be heard in one ear and person B in the other. A number of existing systems are based on this idea. The inventor has taken it several steps further, including gain shifting, distinctive voice-over audio signatures, and automatic microphone switching.—GLA

6,301,572

43.40.Le NEURAL NETWORK BASED ANALYSIS SYSTEM FOR VIBRATION ANALYSIS AND CONDITION MONITORING

Gregory A. Harrison, assignor to Lockheed Martin Corporation
9 October 2001 (Class 706/52); filed 2 December 1998

In this system for long-term condition monitoring of machines, such as gas turbines, the time-domain output of a vibration sensor is transformed to the frequency domain using fast Fourier transform processing. The resulting information is provided to a ''fuzzy adaptive resonance theory'' neural network, which can detect patterns in a signal and which can be trained to recognize signal characteristics or trends. The output can be presented on display devices for presentation to an operator and can be made available for other control and information purposes.—EEU

6,304,654

43.38.Si TELEPHONE HANDSET INTERFACE WITH FEEDBACK CONTROL

Brian Albert Wittman, assignor to Lucent Technologies, Incorporated
16 October 2001 (Class 379/387.01); filed 6 February 1996

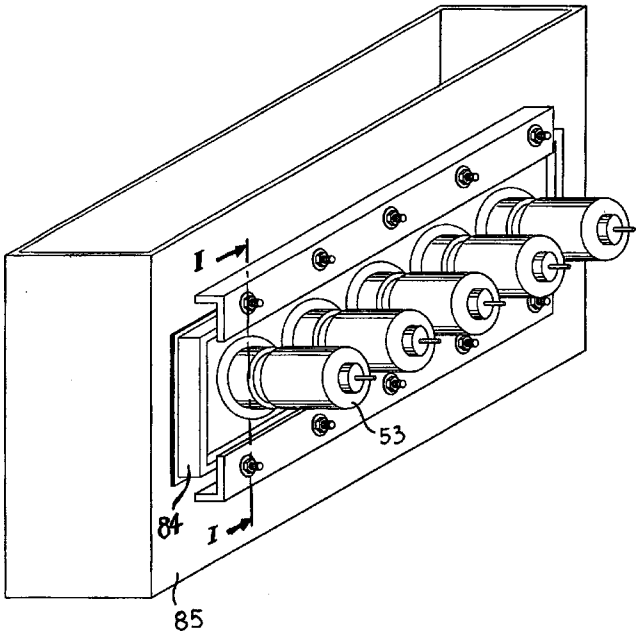
Loss or gain are selectably applied to the receiving and the transmitting portions of a telephone handset to reduce feedback problems. Accordingly, the microphone signal may have a loss applied when the user is listening at high receiver/speaker gain levels. The loss applied to the microphone signal is then reduced when the user begins to speak.—IMH

6,308,585

43.40.Le METHOD AND A DEVICE FOR ATTACHING ULTRASONIC TRANSDUCERS

Bo Nilsson and Håkan Dahlberg, assignors to Ultra Sonus AB
30 October 2001 (Class 73/866.5); filed 10 February 2000

The patent describes a way of attaching transducers **53** to a wall



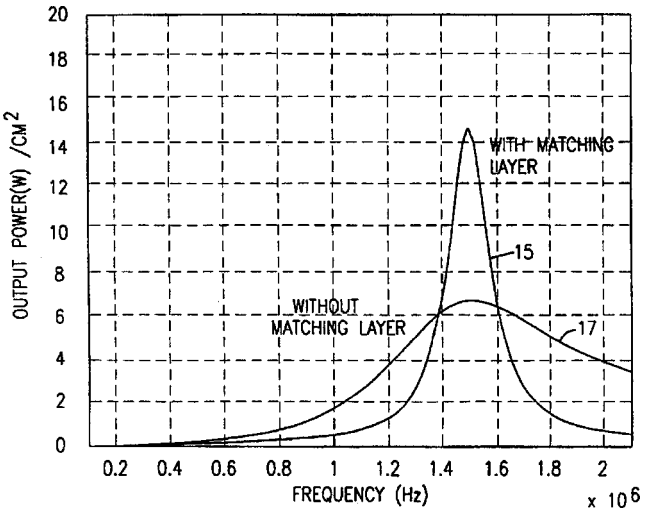
(e.g., vessel wall **85**) by threading the transducers onto a matching threaded coupler **84** affixed to the wall.—IMH

6,307,302

43.40.Rj ULTRASONIC TRANSDUCER HAVING IMPEDANCE MATCHING LAYER

Minoru Toda, assignor to Measurement Specialties, Incorporated
23 October 2001 (Class 310/334); filed 23 July 1999

A transducer coupled to an impedance matching layer is described. The impedance of the matching layer is less than that of the radiation



medium but greater than that of the transducer at the resonance frequency. The result is a more efficient radiation characteristic.—IMH

6,295,871

43.40.Yq VIBRATION SENSING DEVICE

Jung-Tsung Wei, Tainan City, Taiwan, Province of China
2 October 2001 (Class 73/570); filed 30 May 2000

A small electrically conductive ball is suspended inside a conductive spherical shell via a conductive thread that is attached at the tip of a rod, which tip is at the center of the spherical shell. A second conductive ball is free to roll inside the shell. The length of the thread and the diameters of the balls are such that the balls just make contact with each other when the spherical shell is undisturbed, so that electrical continuity is maintained. Motion of the shell causes the balls to lose contact, interrupting the electrical circuit. The spherical geometry apparently permits this sensor to work over a wide range of angles from the vertical.—EEU

6,305,226

43.58.Ls METHOD AND APPARATUS FOR IMAGING ACOUSTIC FIELDS IN HIGH-FREQUENCY ACOUSTIC RESONATORS

Bradley Paul Barber *et al.*, assignors to Agere Systems Guardian Corporation
23 October 2001 (Class 73/606); filed 30 August 1999

This apparatus is capable of measuring sub-Angstrom scale vibrations and discerning sub-nanometer surface features of a specimen. The specimen is vibrated using a rf signal frequency modulated by a second lower frequency signal. A sensing tip, coupled to a control circuit, is used to image vibration modes of the specimen, for example, a piezoelectric specimen.—IMH

6,292,571

43.66.Ts HEARING AID DIGITAL FILTER

Walter P. Sjursen, assignor to Sarnoff Corporation
18 September 2001 (Class 381/312); filed 2 June 1999

A reduced hardware FIR digital filter scheme is described in which multiplies are accomplished with a single general purpose multiplier that uses only shifting. The invention is possible because the digital filter coefficients are related generally to 2^n. Advantages over traditional hardware multipliers include less power consumption, reduced cost, and smaller size without sacrificing performance.—DAP

6,292,572

43.66.Ts HEARING AIDS WITH STANDARDIZED SPHEROIDAL HOUSINGS

Robert Yoest *et al.*, assignors to Beltone Electronics Corporation
18 September 2001 (Class 381/322); filed 19 September 1996

A standard, one-size-fits-all, symmetrical, egg-shaped or pear-shaped housing has a soft, deformable spongelike outer layer to ease insertion and improve comfort for the hearing aid wearer. The soft outer layer may optionally be removable and may contain a wax guard.—DAP

6,307,944

43.66.Ts SYSTEM FOR MITIGATING RF INTERFERENCE IN A HEARING AID

Reginald G. Garratt and Elmer V. Carlson, assignors to Knowles Electronics LLC
23 October 2001 (Class 381/312); filed 2 March 1998

This comb filter is designed to suppress the harmonics of the fixed on–off gating frequency of the rf signal emanating from a digital cellular telephone or another similarly interfering device. The filter is inserted between the hearing aid amplifier and the receiver. To make the attenuated frequencies coincide with the frequencies of the interference signals, the method uses two pulse detectors separated by a delay. The delay is adapted to the proper value to correspond with the time of arrival of the pulses. The delayed pulsed signals are subtracted from the original pulsed signals to produce interference cancellation.—DAP

6,307,945

43.66.Ts RADIO-BASED HEARING AID SYSTEM

Andrew James Jamieson Hall, assignor to Sense-Sonic Limited
23 October 2001 (Class 381/315); filed in the United Kingdom 21 December 1990

A switchable, directional/omnidirectional microphone, a direct electrical input, and an FM transmitter are housed in a small, portable hand-held unit. An FM receiver connected to an amplifier which drives an inductive loop is packaged in a combination pendant/necklace. The inductive signal is coupled to an earpiece which provides an audio output to the ear.—DAP

6,292,769

43.70.Kv SYSTEM FOR AUTOMATED TRANSLATION OF SPEECH

Mary A. Flanagan *et al.*, assignors to America Online, Incorporated
18 September 2001 (Class 704/3); filed 14 February 1995

If you accept the premise that unconstrained speech recognition and language translation are in fact possible, then the ideas covered by this patent are simple. A typical, commercial speech recognizer converts incoming voice to text. The text is automatically translated to multiple languages as requested, and the result is transmitted and resynthesized. Two references allegedly documenting available translation software are in fact hyped up consumer reports, one titled ''Machine Assisted Translation''—hardly an automated translation technique.—DLR

6,305,942

43.71.Hw METHOD AND APPARATUS FOR INCREASED LANGUAGE FLUENCY THROUGH INTERACTIVE COMPREHENSION, RECOGNITION AND GENERATION OF SOUNDS, WORDS AND SENTENCES

Robert S. Block *et al.*, assignors to MetaLearning Systems, Incorporated
23 October 2001 (Class 434/156); filed 12 November 1998

This language teaching software displays an object and speaks the word for that object in any of several languages. In addition, the word may be spelled out on the screen. There is no provision for an obvious extension which would allow the user to speak into the machine and have the voice analyzed for pronunciation accuracy.—DLR

6,302,697

43.71.Ky METHOD AND DEVICE FOR ENHANCING THE RECOGNITION OF SPEECH AMONG SPEECH-IMPAIRED INDIVIDUALS

Paula Anne Tallal, San Francisco, California *et al.*
16 October 2001 (Class 434/185); filed 8 December 1994

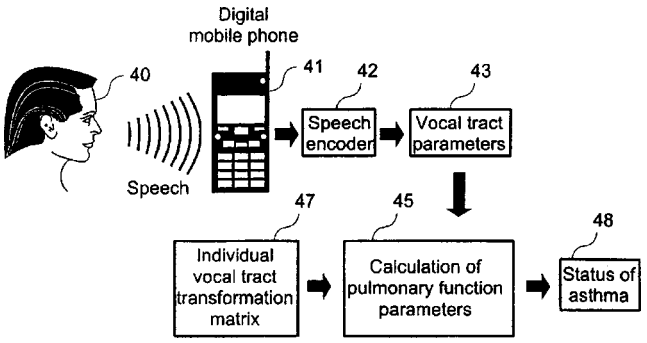
Like many others in a long series of patents related to tests for hearing deficiencies, this patent describes a method of stretching the durations of those portions of the speech signal which involve rapid spectral transitions. The current system also produces an increase in the signal amplitude during these intervals. In addition to the therapeutic value, the techniques are said to be of value in learning the sounds of a foreign language.—DLR

6,289,313

43.72.Ar METHOD, DEVICE AND SYSTEM FOR ESTIMATING THE CONDITION OF A USER

Pekka Heinonen *et al.*, assignors to Nokia Mobile Phones Limited
11 September 2001 (Class 704/270); filed in Finland 30 June 1998

This voice analyzer system would construct parameters from linear prediction encoded data packets to determine speaker conditions, such as excessive breathing, agitation, fatigue, etc. The measurements could be used



to monitor for unhealthy conditions, such as asthma. Based on standard cell phone speech coding systems, the method would not need a special transmitting device.—DLR

6,308,153

43.72.Fx SYSTEM FOR VOICE VERIFICATION USING MATCHED FRAMES

William Y. Huang *et al.*, assignors to ITT Defense, Incorporated
23 October 2001 (Class 704/246); filed 10 April 1996

This speaker verification system collects the speech from a known phrase spoken by the applicant and performs a typical HMM recognition process to assign phonetic labels to the speech frames. Tests for energy range and spectral equalization allow some of the test frames to be eliminated. The applicant's labeled frames are then compared with reference frames having the same phonetic labels.—DLR

6,289,311

43.72.Ja SOUND SYNTHESIZING METHOD AND APPARATUS, AND SOUND BAND EXPANDING METHOD AND APPARATUS

Shiro Omori and Masayuki Nishiguchi, assignors to Sony Corporation
11 September 2001 (Class 704/268); filed in Japan 23 October 1997

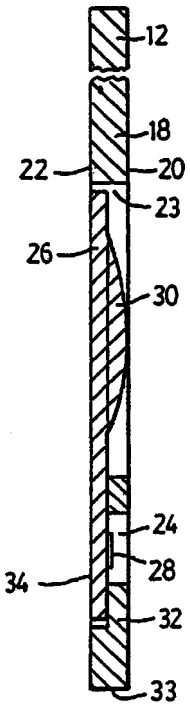
This linear prediction speech coder with codebooks for both voiced and unvoiced speech uses two sets of codebooks, one set for narrow-band, low-bit-rate synthesis and one set for wideband, high-bit-rate, high-quality synthesis. The object of the invention can be stated fairly simply: speech codes, encoded and transmitted with the low-bit-rate system, could be reproduced with the high-quality system. The objection can be stated even more simply. If it's so easy to produce high-quality speech from low-bit-rate codes, why bother with a low-bit-rate synthesizer?—DLR

6,292,780

43.72.Ja TALKING TRADING CARD PLAYER SYSTEM

Dieter D. Doederlein *et al.*, assignors to Micra SoundCards, Incorporated
18 September 2001 (Class 704/270); filed 25 August 1995

This patent describes an enhanced type of collectable trading card, such as baseball cards or hockey cards. Within the cardboard card would be a synthesizer chip containing audio samples of speech or sounds appropriate



to the card's content. The card would be placed into a reader/player unit, which would supply power and a loudspeaker to hear the card's output.—DLR

6,292,777

43.72.Kb PHASE QUANTIZATION METHOD AND APPARATUS

Akira Inoue and Masayuki Nishiguchi, assignors to Sony Corporation
18 September 2001 (Class 704/230); filed in Japan 6 February 1998

This is a fairly typical line-spectral-pairs vocoder using code-excited linear prediction for the speech analysis/synthesis. But, in addition to the usual CELP codebook parameters, the system also transmits phase information to allow correct reconstruction of the harmonics' phases during voiced segments. The encoded speech quality is said to be improved as a result.—DLR

6,295,520

43.72.Kb MULTI-PULSE SYNTHESIS SIMPLIFICATION IN ANALYSIS-BY-SYNTHESIS CODERS

Wenshun Tian, assignor to Tritech Microelectronics Limited
25 September 2001 (Class 704/223); filed 15 March 1999

This patent describes an enhancement to the G.723.1 CELP vocoder standard in which a speedup in the computation of vocal tract response functions is obtained by sorting the excitation signal into zero and nonzero values. Only the nonzero values are processed by the tract convolution filter.—DLR

6,308,798

43.80.Ev LIGHTWEIGHT STETHOSCOPE WITH VARIABLE DIAPHRAGM AND BELL COMPONENTS

Richard Rashman and Dennis Shick, assignors to Prestige Medical Corporation
30 October 2001 (Class 181/131); filed 22 April 1996

This is yet another stethoscope design, this one featuring a dual-head chest piece with removable diaphragm structures and bell components for adapting the stethoscope for use with infants or adults, or on sites having less skin area, such as ribs. The stethoscope also incorporates a single tube for connecting the chest piece to the earpiece. The unit is said to be smaller and lighter in weight than a conventional model.—DRR

6,308,714

43.80.Gx ULTRASOUND ENHANCED CHEMOTHERAPY

Thomas M. Peterson and Robert J. Siegel, assignors to Coraje, Incorporated
30 October 2001 (Class 128/898); filed 10 November 1998

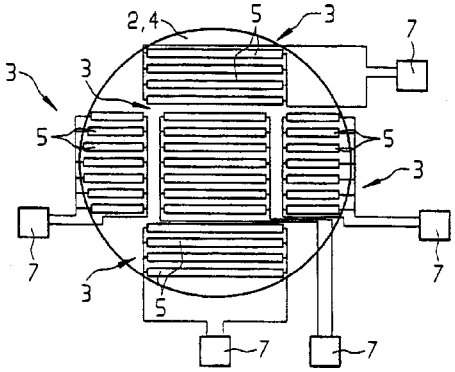
This apparatus is designed to enhance the action of anti-cancer agents. The anti-cancer agent is introduced into or near a tumor and then ultrasonic energy is directed at the tumor. The ultrasonic energy is claimed to be sufficient to increase the anti-cancer activity on the solid tumor without significant heating of the tumor or surrounding tissue.—DRR

6,307,303

43.80.Vj ULTRASOUND TRANSMITTING CONFIGURATION

Mario Bechtold *et al.*, assignors to Siemens Aktiengesellschaft
23 October 2001 (Class 310/335); filed in Germany 23 July 1998

An ultrasonic transducer arrangement and lens system is shown. Banks of elongated transducer elements are arranged behind an acoustic lens. Each bank is driven as a unit, and the depth of focus of the overall device is adjustable by choosing to drive one bank at the center of the lens, or by choosing to also drive several banks arranged around the center of the lens.—IMH



6,305,225

43.80.Vj ULTRASONIC SIGNAL FOCUSING METHOD FOR ULTRASONIC IMAGING SYSTEM

Moo-Ho Bae and Mok-Kun Jeong, assignors to Medison Company, Limited
23 October 2001 (Class 73/602); filed in the Republic of Korea 9 December 1998

A method is described for use in multi-element ultrasonic imaging systems, wherein the element-to-focus distance for each element is estimated, then refined, based on the image brightness or contrast. The method yields a best focusing time delay curve for the elements, especially in media (e.g., biological) where adjacent segments (of tissue) may have different sound speeds.—IMH

A simple formula for the Lamb modes in a plate (L)

Faiz Ahmad[a)]
Department of Mathematics, Quaid-I-Azam University, Islamabad, Pakistan

(Received 15 April 2001; revised 17 January 2002; accepted 17 January 2002)

The dispersion relation for the Rayleigh–Lamb modes in an elastic plate can be replaced by a simpler equation, which admits exact solution. The approximation is valid for almost all modes when the phase velocity is between c_T and c_L. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1460921]

PACS numbers: 43.20.Bi, 43.20.Mv [ANN]

I. INTRODUCTION

Consider an infinite isotropic plate of thickness $2h$ characterized by the phase speeds c_T and c_L, respectively, of the transverse and longitudinal bulk waves. Let ω and k denote the frequency and the wave number of a wave which propagates in a direction parallel to the plate surfaces which are assumed to be free of traction. The dispersion relation for the *symmetric* modes is given by[1]

$$\frac{\tan(qh)}{\tan(ph)} = \frac{-4pqk^2}{(q^2-k^2)^2}, \tag{1}$$

where

$$p = \sqrt{\frac{\omega^2}{c_L^2} - k^2}, \tag{2}$$

$$q = \sqrt{\frac{\omega^2}{c_T^2} - k^2}. \tag{3}$$

The corresponding dispersion relation for the antisymmetric modes is obtained from Eq. (1) by interchanging p and q on the left hand side of the equation. In this letter we shall discuss only symmetric modes.

Equation (1) is known as the Rayleigh–Lamb equation.[2,3] Dispersion curves, expressing ω or the phase speed in terms of the wave number, are obtained numerically. Sketching of these curves is facilitated, to a great extent, by Mindlin's method of bounds.[4] Details of these curves may be found in virtually any graduate textbook on elastic waves, see, for example, Ref. 1 or 5.

In this letter we shall show that, for $c_T < c < c_L$, Eq. (1) can be replaced by a simpler equation which admits an exact solution. This makes the sketching of modes a simple affair. If the spectrum is plotted in the form of the phase velocity as a function of the dimensionless wave number, a horizontal line intersects the modes at equidistant points.

II. THE APPROXIMATE EQUATION

Define

$$c = \frac{\omega}{k}, \quad u = hk, \tag{4}$$

and assume $c_T < c < c_L$. We can write Eq. (1) in the form

$$\frac{\tan(u\alpha)}{\tanh(u\beta)} = \gamma, \tag{5}$$

where

$$\alpha(c) = \sqrt{\frac{c^2}{c_T^2} - 1}, \tag{6}$$

$$\beta(c) = \sqrt{1 - \frac{c^2}{c_L^2}}, \tag{7}$$

and

$$\gamma(c) = \frac{4\alpha(c)\beta(c)}{(1-\alpha(c)^2)^2}. \tag{8}$$

The function $\tanh x$ is an increasing function bounded above by unity. For any $u\beta > 0$, $1-\varepsilon < \tanh(u\beta) < 1$, and ε becomes progressively small as $u\beta$ becomes large. For example, if

$$u\beta > 3, \tag{9}$$

then $\varepsilon < 0.005$, and the approximation

$$\tanh(u\beta) \cong 1 \tag{10}$$

can be used with an accuracy of 5 parts in 1000. Since condition (9) is satisfied for almost all modes s_i, $i \geqslant 2$, the approximation (10) is valid in most parts of the region $c_T < c < c_L$. Equation (5) is now replaced by

$$\tan(\alpha u) = \gamma, \tag{11}$$

with the solution

$$u_n = \frac{\tan^{-1}(\gamma) + n\pi}{\alpha}, \quad n = 1,2,3,\ldots. \tag{12}$$

Equation (12) is an explicit solution of the Rayleigh–Lamb equation and can be used to plot the dispersion curves s_2, s_3,…., with fairly good accuracy in the region where Eq. (10) is valid. Note that the points given by Eq. (12) are such that the distance between a pair of adjacent points is constant along a horizontal line. Thus if any one of the dispersion curves is sketched, all other modes can be obtained by displacing each point horizontally by π/α units. Note that the value of π/α differs along different lines.

[a)]Electronic mail: faizmath@hotmail.com

TABLE I. Dimensionless wave number for various modes for an aluminum plate, Poisson's ratio $\nu=0.355$, corresponding to the phase velocity $c=1.8c_T$.

Modes	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
Approximate Eq. (12)	2.844	4.943	7.042	9.141	11.240	13.338	15.438	17.537
Exact Eq. (5)	2.818	4.940	7.041	9.141	11.240	13.338	15.438	17.537

III. VALIDITY OF THE APPROXIMATION

If $\tanh(u\beta)=1-\varepsilon$, then Eq. (5) gives

$$\tan(u\alpha)=\gamma(1-\varepsilon),$$

or

$$u_n=\frac{\tan^{-1}(\gamma(1-\varepsilon))+n\pi}{\alpha}. \tag{13}$$

Thus the error du in the value of u, by using Eq. (12) instead of Eq. (13), is

$$du\cong\frac{\gamma\varepsilon}{\alpha(1+\gamma^2)}=\frac{4\beta\varepsilon}{([\alpha(c)]^2-1)^2(1+\gamma^2)}. \tag{14}$$

Since ε becomes very small for the modes $s_2, s_3, \ldots$ for practically the entire range $c_T<c<c_L$ and even for s_1, for nearly three-quarters of the range, the approximation (12) gives very good results. In Table I, we compare the exact results with those given by Eq. (12) for aluminum when $c=1.8c_T$. We see that for s_1 the difference between the two results is less than 1% and for s_2 it is less than 4 parts in 5000. For the rest of the modes the two results are essentially identical.

In Fig. 1 are plotted the symmetric modes s_1, s_2, s_3 and s_4. The approximate curves obtained by using Eq. (12) are shown dashed in the figure. It is found that the two types of curves differ significantly only for s_1 between $1.8c_T$ and c_L. The curves for the higher modes essentially overlap throughout the entire range $c_T<c<c_L$, the difference being so small as to be invisible on the scale used for the figure.

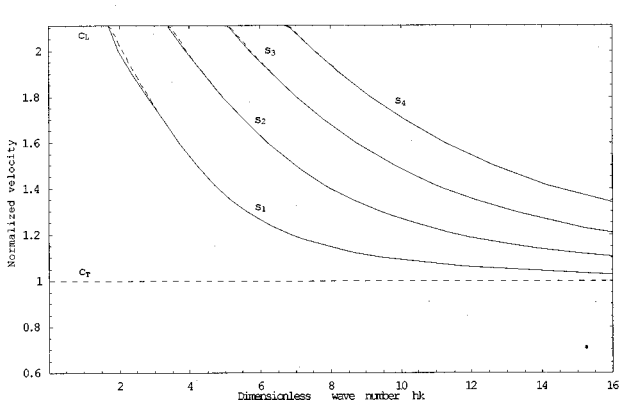


FIG. 1. Symmetric modes s_1-s_4 in an aluminum plate. Full curves are exact and dashed curves are based on Eq. (12).

ACKNOWLEDGMENTS

Financial support for this work was provided by the University Research Fund of the Quaid-i-Azam University. The author is also grateful to Moniba Shams for her help in the preparation of the manuscript.

[1] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1980).
[2] Lord Rayleigh, "On the free vibrations of an infinite plate of homogeneous isotropic elastic matter," Proc. London Math. Soc. **20**, 225–235 (1889).
[3] H. Lamb, "On the flexure of an elastic plate," Proc. London Math. Soc. **21**, 70–90 (1889).
[4] R. D. Mindlin, "Waves and vibrations in isotropic elastic plates," in *First Symposium on Structural Mechanics, 1958*, edited by J. N. Goodier and N. J. Hoff (Pergamon, Oxford, 1960), pp. 199–232.
[5] J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge U. P., Cambridge, 1999).

Chaotic behavior of piezoelectric plate vibration (L)

Yu Zhang,[a] Wen-hua Jiang, and Gong-huan Du
Institute of Acoustics, State Key Laboratory of Modern Acoustics, Nanjing University, Nanjing 210093

(Received 11 November 1997; revised 6 December 2000; accepted 14 February 2002)

A dynamical model for the piezoelectric plate vibrations is presented in this letter. The calculations of phase portrait and Lyapunov exponents based on the coupled-mode equations given by the theoretical model show that the vibrations of a piezoelectric plate resonator under the resonance may exhibit the chaotic behavior. The experiment is carried out for a $LiNbO_3$ thickness-longitudinal vibration plate resonator by using a setup based on heterodyne laser probe. The fractional harmonic vibration as well as chaotic behaviors are observed. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1471900]

PACS numbers: 43.25.Gf, 43.25.Rq [MAB]

I. INTRODUCTION

The linear vibration of a piezoelectric plate has been thoroughly studied.[1] Recently the nonlinear behaviors were interesting to many authors.[2–4] Fractional harmonic vibrations had been observed in a resonance ultrasonic wave system[5] as well as in an ultrasonic Langevin transducer.[6] When the nonlinear phenomena are involved, the perturbation method of successive approximation is commonly used.[4] However, it is known that the successive approximation theory of straightforward expansion often leads to appearance of secular terms and is not adequate to describe the nonlinear behaviors shown by a piezoelectric plate resonator under resonant driving. In this letter a new dynamical model for piezoelectric plate vibrations is presented. First, the theoretical derivation to obtain the coupled-mode dynamical equations is given. In the derivation, the method of boundary integration and modal truncation is used. As a result, the partial differential equation governing the piezoelectric plate vibration is reduced to a set of ordinary differential equations. Then the nonlinear vibrations of a piezoelectric plate resonator are analyzed based on the equations. The phase portrait and Lyapunov exponents are calculated. Finally, the experiment for a $LiNbO_3$ plate resonator is carried out. The fractional harmonic vibration and chaotic behavior are experimentally observed. For the piezoelectric plate vibration under electric driving, it is first calculated in theory and observed in experiment.

II. DYNAMICAL MODEL AND ANALYSIS OF PIEZOELECTRIC PLATE VIBRATION

The calculation is conducted for a purely thickness-longitudinal vibration piezoelectric plate. The geometry of calculation is shown in Fig. 1. The surfaces of the plate are electroded through which a sinusoidal driving voltage of the plate resonator is applied. The coordinate axis x_3 is parallel to the normal of the plate and corresponds to a pure mode direction of the piezoelectric crystal. The dimensions along x_1 and x_2 are considered infinite. Thus this is a one-dimensional problem. Then the equations governing the plate vibration are

$$\rho_0 \frac{\partial^2 u_3}{\partial t^2} = \frac{\partial P_{33}}{\partial x_3}, \quad \frac{\partial D_3}{\partial x_3} = 0, \tag{1}$$

where u_3 is the particle displacement, P_{33} is the Piola–Kirchoff stress, and D_3 is the electric displacement. The boundary conditions are expressed as

$$P_{33}|_{x_3=\pm h/2} = 0, \quad \int_{-h/2}^{h/2} E_3\, dx_3 = V_0 \cos \Omega t, \tag{2}$$

where h is the thickness of the plate, V_0 and Ω are amplitude and angular frequency of driving voltage, respectively, and E_3 is electric field. Keeping up to cubic of $\partial u_3/\partial x_3$ and linear terms of D_3, the constitute equation[7] is yielded that

$$P_{33} = C_{33}^D \frac{\partial u_3}{\partial x_3} + \alpha \frac{\partial^2 u_3}{\partial x_3 \partial t} - \frac{e_{33}}{\varepsilon_{33}} D_3 + \frac{1}{2}(\bar{C}_{333} + 3C_{33}^D)\left(\frac{\partial u_3}{\partial x_3}\right)^2 - \frac{e_{33} - \bar{d}_{3\cdot 33}}{\varepsilon_{33}} \frac{\partial u_3}{\partial x_3} D_3 + \frac{1}{6}(\bar{C}_{3333} + 6\bar{C}_{333} + 3C_{33}^D) \times \left(\frac{\partial u_3}{\partial x_3}\right)^3, \tag{3a}$$

$$D_3 = \varepsilon_{33} E_3 + e_{33}(\partial u_3/\partial x_3), \tag{3b}$$

where the stiffened second-order elastic constant, the effective third-order elastic constant, the effective first odd electroelastic constant, and the effective fourth-order elastic constant, respectively, are $C_{33}^D = C_{33} + e_{33}^2/\varepsilon_{33}$, $\bar{C}_{333} = C_{333} + \varepsilon_{333}(e_{33}/\varepsilon_{33})^3 - 3b_{33}(e_{33}/\varepsilon_{33})^2 - 3d_{3\cdot 33}(e_{33}/\varepsilon_{33})$, $\bar{d}_{3\cdot 33} = d_{3\cdot 33} - \varepsilon_{333}(e_{33}/\varepsilon_{33})^2 + 2b_{33}(e_{33}/\varepsilon_{33})$, $\bar{C}_{3333} = C_{3333} + 3(\varepsilon_{333}^2/e_{33})(e_{33}/\varepsilon_{33})^5 - (\varepsilon_{3333} + 12b_{33}\varepsilon_{333}/e_{33})(e_{33}/\varepsilon_{33})^4 + (4d_{333\cdot 3} + 12b_{33}^2/e_{33} - 6d_{3\cdot 33}\varepsilon_{333}/e_{33})(e_{33}/\varepsilon_{33})^3 + 6(a_{3\cdot 33} + 2d_{3\cdot 33}b_{33}/e_{33})(e_{33}/\varepsilon_{33})^2 - (4d_{33\cdot 33} - 3d_{3\cdot 33}^2/e_{33})(e_{33}/\varepsilon_{33})$. In the expressions, C_{33}, e_{33}, and ε_{33} are the second-order elastic, piezoelectric and dielectric constants, respectively. C_{333} and C_{3333} are the third- and fourth-order elastic constants. ε_{333} and ε_{3333} are the third- and fourth-order dielectric constants. b_{33} is the electrostrictive constant.

[a]Current address: Department of Surgery, Division of Otolaryngology Head and Neck Surgery, University of Wisconsin Medical School, Madison, WI 53792-7375. Electronic mail: zhang@surgery.wisc.edu

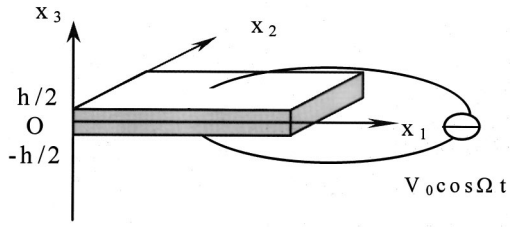


FIG. 1. The geometry of the calculation.

$d_{3\cdot 33}$, $d_{33\cdot 33}$, $d_{333\cdot 3}$, and $a_{3\cdot 33}$ are the first odd, second odd, third odd, and first even electroelastic constants, respectively. α is viscosity constant. Substituting these expressions, the equation of motion is reduced to

$$Lu_3=\rho_0\frac{\partial^2 u_3}{\partial t^2}-C_{33}^D\frac{\partial^2 u_3}{\partial x_3^2}-\alpha\frac{\partial^3 u_3}{\partial x_3^2\partial t}$$
$$=\frac{\partial}{\partial x_3}\left[\frac{1}{2}(\bar{C}_{333}+3C_{33}^D)\left(\frac{\partial u_3}{\partial x_3}\right)^2-\frac{e_{33}-\bar{d}_{3\cdot 33}}{\varepsilon_{33}}D_3\frac{\partial u_3}{\partial x_3}\right.$$
$$\left.+\frac{1}{6}(\bar{C}_{3333}+6\bar{C}_{333}+3C_{33}^D)\left(\frac{\partial u_3}{\partial x_3}\right)^3\right]. \tag{4}$$

It is known that the governing equation and boundary condition for a freely vibrating linear plate are expressed as[1]

$$L'\nu_3=\rho_0\frac{\partial^2\nu_3}{\partial t^2}-C_{33}^D\frac{\partial^2\nu_3}{\partial x_3^2}=0, \tag{5a}$$

$$\partial\nu_3/\partial x_3|_{x_3=\pm h/2}=0. \tag{5b}$$

If Eq. (5) is taken as the adjoint equation, combining Eq. (4) with Eq. (5) and integrating along the thickness of the plate yields

$$\int_{-h/2}^{+h/2}(\nu_3 Lu_3-u_3L'\nu_3)\,dx_3$$
$$=\int_{-h/2}^{+h/2}\left[\rho_0\left(\nu_3\frac{\partial^2 u_3}{\partial t^2}-u_3\frac{\partial^2\nu_3}{\partial t^2}\right)\right.$$
$$\left.-C_{33}^D\left(\nu_3\frac{\partial^2 u_3}{\partial x_3^2}-u_3\frac{\partial^2\nu_3}{\partial x_3^2}\right)-\alpha\nu_3\frac{\partial^3 u_3}{\partial x_3^2\partial t}\right]dx_3$$
$$=\int_{-h/2}^{+h/2}\nu_3\frac{\partial}{\partial x_3}\left[\frac{1}{2}(\bar{C}_{333}+3C_{33}^D)\left(\frac{\partial u_3}{\partial x_3}\right)^2\right.$$
$$-\frac{e_{33}-\bar{d}_{3\cdot 33}}{\varepsilon_{33}}D_3\frac{\partial u_3}{\partial x_3}$$
$$\left.+\frac{1}{6}(\bar{C}_{3333}+6\bar{C}_{333}+3C_{33}^D)\left(\frac{\partial u_3}{\partial x_3}\right)^3\right]dx_3. \tag{6}$$

Here the boundary condition equations (2) and (5b) and constitutive equation (3) are used. The solution of Eq. (5) can be expressed as

$$\nu_3=\sin\omega_0 t\sin k_0 x_3+\sin 2\omega_0 t\cos 2k_0 x_3+\cdots, \tag{7}$$

where k_0 and ω_0 are wave number and angular frequency of the fundamental eigenmode and $k_0 h=\pi$. The particle displacement u_3 is assumed to take the form of

$$u_3=A_1(t)\sin k_0 x_3+A_2(t)\cos 2k_0 x_3+\cdots, \tag{8}$$

where $A_1(t)$ corresponds to an antisymmetric mode and $A_2(t)$ to a symmetric mode and they are functions with respect to time t. Since the electromechanical nonlinearities in constitutive equation are ignored for simplicity, Eq. (2) easily leads to

$$D_3 h=\varepsilon_{33}V_0\cos\Omega t+e_{33}u_3|_{-h/2}^{+h/2}. \tag{9}$$

Substituting (7)–(9) into (6) and implementing integration under the condition that the modes are truncated at the first two modes, a set of nondimensional coupled-mode equations which describe the dynamical behaviors of the piezoelectric plate vibration is given by

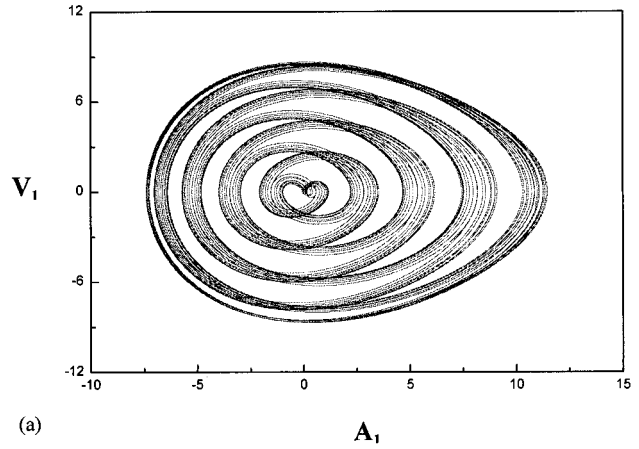
$$\ddot{\tilde{A}}_1+\omega_1^2\tilde{A}_1=\frac{4}{\pi^2}\tilde{V}_0\cos\tilde{\Omega}\tilde{t}+\Gamma_1\frac{2\tilde{A}_1^2}{\pi}-\frac{64}{15}\gamma_{cc}\frac{\tilde{A}_2^2}{\pi}$$
$$+\gamma_{ee}\frac{\tilde{A}_1}{\pi}\tilde{V}_0\cos\tilde{\Omega}\tilde{t}+\gamma_{ccc}\left(\frac{\tilde{A}_1^3}{8}+\tilde{A}_1\tilde{A}_2^2\right)-\tilde{\alpha}\dot{\tilde{A}}_1, \tag{10a}$$

$$\ddot{\tilde{A}}_2+\omega_2^2\tilde{A}_2=\Gamma_2\frac{8\tilde{A}_1\tilde{A}_2}{\pi}+\gamma_{ee}\frac{4\tilde{A}_2}{\pi}\tilde{V}_0\cos\tilde{\Omega}\tilde{t}$$
$$+\gamma_{ccc}(\tilde{A}_1^2\tilde{A}_2+2\tilde{A}_2^3)-4\tilde{\alpha}\dot{\tilde{A}}_2, \tag{10b}$$

where the dot over a letter denotes time derivative. $\tilde{A}_1=k_0A_1$, $\tilde{A}_2=k_0A_2$, $\tilde{t}=\omega_0 t$, $\tilde{\Omega}=\Omega/\omega_0$, $\tilde{V}_0=\gamma_{ec}k_0V_0$, $\omega_1^2=(1-8k_t^2/\pi^2)$, and $\omega_2^2=4$ are normalized vibration amplitude, time, frequency and driving voltage, respectively, and $\gamma_{ec}=e_{33}/C_{33}^D$, $\gamma_{cc}=(3C_{33}^D+\bar{C}_{333}/C_{33}^D)$, $\gamma_{ee}=(e_{33}-d_{3\cdot 33})/e_{33}$, $\gamma_{ccc}=-(C_{3333}+6C_{333}+3C_{33}^D)/C_{33}^D$, $\Gamma_1=-\frac{2}{3}\gamma_{cc}+k_t^2\gamma_{ee}$, $\Gamma_2=-\frac{16}{15}\gamma_{cc}+k_t^2\gamma_{ee}$, $\tilde{\alpha}=(\alpha/C_{33}^D)\omega_0$, where k_t^2 is electromechanical coupling coefficient.

Now the nonlinear behaviors of the piezoelectric plate vibrations are analyzed based on Eq. (10). As an example, the analysis is conducted for a Z-cut $LiNbO_3$ plate resonator. In the analysis, the following material parameters for $LiNbO_3$ are used: $k_t^2=0.026$, $\gamma_{cc}=-1.3\times10^5$, $\gamma_{ee}=-6.692$, $\Gamma_1=8.667\times10^4$, $\Gamma_2=1.387\times10^5$, $\alpha=1.669\times10^{-6}$, $\omega_1=0.989$, and $\omega_2=2$. Here the value of γ_{cc} is determined by the experiment described in Ref. 4. In the calculation, it is obtained by the measurement under the resonance frequency. We find that the influence of the value of γ_{ccc} on dynamics is not significant. Since the value of γ_{ccc} cannot be found in literature, the value is plausibly taken as -6.6×10^5.

When the driving frequency $\Omega=1.001$ is near to primary resonance frequency 1, the antisymmetric primary resonance mode A_1 will be strongly excited. If the driving is strong enough, the A_2 mode will be parametrically excited. When the normalized driving voltage $V_0=1.6\times10^{-6}$ (or the voltage is around 75 V peak-peak), the vibration exhibits chaotic state. The phase portraits are shown in Figs. 2(a) and (b). The characteristic of chaos is evidently observed. The Lyapunov spectra are calculated as $\gamma_1=1.735\times10^{-3}$, $\lambda_2=6.97\times10^{-4}$, $\lambda=-5.97\times10^{-4}$, and $\lambda_3=-1.843\times10^{-3}$. The existence of the positive Lyapunov exponents indicates the chaotic state of the piezoelectric plate vibration. One notes that the values of Lyapunov exponents are small. It is



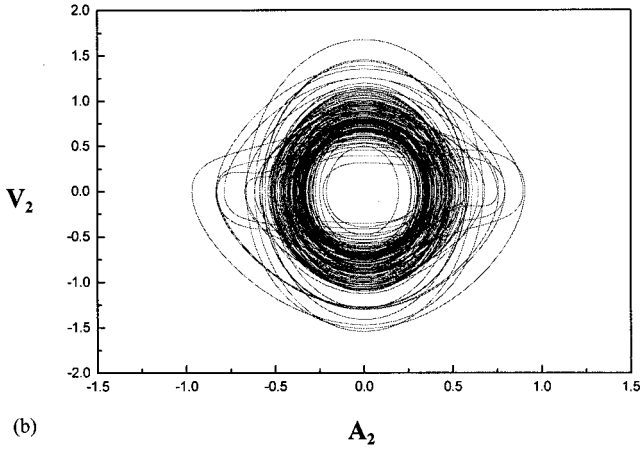


FIG. 2. (a) The calculated phase portraits. Horizontal: vibration displacement amplitude $A_1(2.5\times10^{-10})$. Vertical: vibration velocity amplitude $V_1(2.5\times10^{-10})$. (b) The calculated phase portraits. Horizontal: vibration displacement amplitude $A_2(2.5\times10^{-10})$. Vertical: vibration velocity amplitude $V_2(2.5\times10^{-10})$.

much related to the weak chaotic behavior of the piezoelectric plate vibration under this driving voltage. However, in the post transient regime, the values of λ_1 and λ_2 always keep positive instead of fluctuating between positive and negative values. Therefore we cannot regard them as zero. The system behaves as the weak hyperchaotic system since they have two positive Lyapunov exponents. In order to investigate the evolution of the dynamical behavior of the piezoelectric plate vibration with the increase of driving voltage, we calculate the dependence of the maximal Lyapunov exponent λ_1 on the driving voltage (see Fig. 3). Here, the values of the λ_1 and the voltage V_0 are normalized as $\tilde{\lambda}_1 = \lambda_1/1.735\times10^{-3}$ and $\tilde{V}_0 = V_0/1.6\times10^{-6}$, respectively. It can be seen that when $\tilde{V}_0 < 0.1$, $\tilde{\lambda}_1$ approaches zero. The nonlinear system behaves as the periodic motion. With the increase of voltage V_0 to 0.8 (or the voltage is around 60 V peak-peak), the behaviors of the system become irregular since the maximal positive Lyapunov exponent exists. The vibration of the resonator shows chaos. However, we failed to increase further the driving voltage because we find that

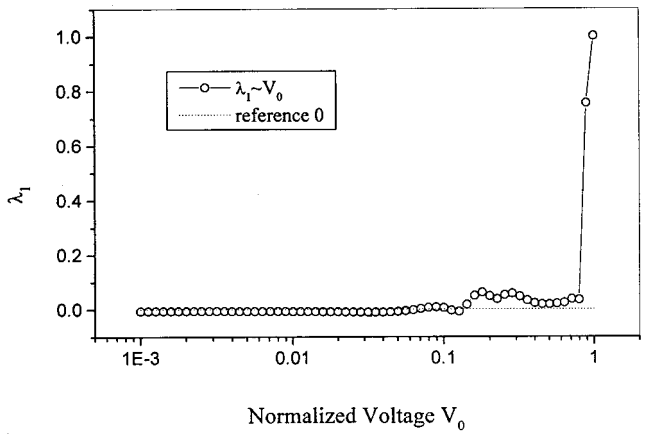


FIG. 3. The dependence of the normalized maximal Lyapunov exponent on the driving voltage.

too high voltage is unpractical to experiment (it is possible to make piezoelectric plate broken). Thus, we obtain the "weak" chaotic solution. It also can be seen in the following experiment.

III. EXPERIMENT

The experiment is carried out for the Z-cut $LiNbO_3$ plate resonator. The setup is shown in Fig. 4(a). To observe the fractional harmonic vibration and chaotic behavior of the piezoelectric plate resonator, a vibration amplitude detector with wide frequency bandwidth is necessary. Here the advantage of a heterodyne laser probe is taken. Since the heterodyne laser probe is a nonresonance amplitude detector, its frequency bandwidth could be very wide. In fact the bandwidth is limited only by the cutoff frequency of the low-pass filter built in the probe. The principle of heterodyne laser probe can be referred to in Ref. 8. With the aid of the laser probe, we can unambiguously detect the fractional harmonic vibration and observe chaotic behavior of the piezoelectric plate resonator through monitoring the occurrence of the lateral sideband or noiselike spectrum between the frequency ω_B of the Bragg cell used in the laser probe and $\omega_B \pm \omega_1$ (ω_1 is the fundamental frequency of the resonator).

The plate resonator used in the experiment is a Z-cut $LiNbO_3$ disc with the thickness of 0.8 mm and diameter of 30 mm. The primary resonance frequency is 4.552 MHz. When the driving frequency Ω is tuned to 4.557 MHz, i.e., a bit higher than its primary resonance frequency, and the driving voltage varies from 30 to 60 V peak-peak, the vibration frequency spectrum of the resonator is shown in Figs. 4(b)–(d). In the figures, the central spectrum is corresponding to the carrier frequency of phase-modulated signal output from the laser probe ($\omega_B = 70$ MHz for the laser probe used in the experiment), and the spectra located on both sides of the central spectrum are the lateral sidebands. The amplitude ratio of the lateral sideband to the carrier frequency (center spectrum) reflects the vibration strength.[8] For example, the lateral sidebands corresponding to $\omega_B \pm \Omega$ is about 20 dB lower than the central spectrum in Fig. 4(b), which means that the vibration amplitude of the resonator is around 10 nm. In Fig. 4(b) only primary resonance (i.e., $\omega_B \pm \omega_1$ and ω_B

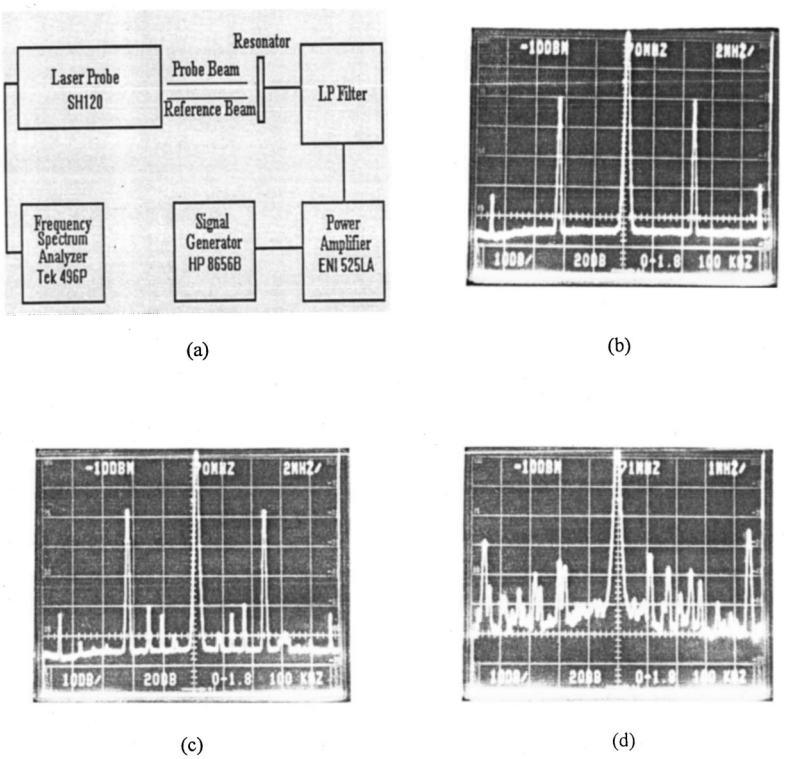


FIG. 4. (a) The experimental setup. (b) The spectrum of resonator oscillation (driving voltage: 30 Vp-p). Horizontal: 2 MHz/div. Vertical: 10 dB/div. (c) The spectrum of resonator oscillation (driving voltage: 40 Vp-p). Horizontal: 2 MHz/div. Vertical: 10 dB/div. (d) The spectrum of resonator oscillation (driving voltage: 60 Vp-p). Horizontal: 1 MHz/div. Vertical: 10 dB/div.

$\pm 2\omega_1$) can be seen. Fractional harmonic vibrations of the resonator occur when driving strength increases, as shown in Fig. 4(c). When the driving strength increases further, the vibrations of the resonator approach chaos, as indicated by Fig. 4(d). Here, we also note that the chaotic output of the experiment is weak. In this case the value of driving voltage (60 Vp-p) is in fairly good agreement with that (75 Vp-p) given by theoretical calculations in Sec. III. If the voltage of the driving is increased further, the piezoelectric plate will break.

IV. CONCLUSIONS

The dynamical model for thickness longitudinal vibrations of a piezoelectric plate is developed by means of boundary integration and modal truncation. The nonlinear behaviors of the piezoelectric plate vibration can be analyzed based on the model. It is seen that the vibration of a piezoelectric plate resonator may evolve from periodic to chaotic motion with the increase of driving strength in the vicinity of primary resonance frequency. The experiment conducted for a $LiNbO_3$ piezoelectric plate resonator demonstrates the fairly good agreement of experiment with theory.

Obviously the occurrence of the chaotic behavior is associated with the extremely strong nonlinearity exhibited by the plate resonator under resonant driving. There are some reasons to make the piezoelectric materials exhibit strong nonlinearity. One example is that extremely strong nonlinearity was observed for PZT ceramics at the Curie temperature.[9] Here the strong nonlinearity of the plate resonator is certainly related to its resonance. The justification will be given in future work.

ACKNOWLEDGMENTS

The authors are grateful to Dr. M. A. Breazeale of National Center of Physical Acoustics, Mississippi, for his instructive discussions. The authors also acknowledge NSF of China for support for this work under Grant No. 19834040.

[1] H. F. Tiersten, *Linear Piezoelectric Plate Vibrations* (Plenum, New York, 1969).
[2] J. J. Gagnepain and R. Besson, "Nonlinear effects in piezoelectric crystals," in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1975), p. 245.
[3] H. F. Tiersten and A. Ballato, "Nonlinear extensional vibrations of quartz rod," J. Acoust. Soc. Am. **73**, 2022 (1983).
[4] W. Jiang, S. Zhou, X. Li, and Y. Shui, "The second harmonic generation in piezoelectric resonators," in Proc. IEEE International Ultrasonics Symposium, Cannes, France, 1994, p. 969.
[5] L. Adler and M. A. Breazeale, "Generation of fractional harmonics in a resonance ultrasonic wave system," J. Acoust. Soc. Am. Suppl. 1 **48**, S1077 (1970).
[6] D. Guyonar, N. Aurelle, C. Richard, P. Gonnard, and L. Eyraud, "Nonlinearities in Langevin transducers," in Proceedings IEEE International Ultrasonics Symposium, Cannes, France (1994), p. 925.
[7] D. A. Berlincourt, D. B. Curran, and H. Jaffe, "Piezoelectric and piezomagnetic materials and their functions in transducers," in *Physical Acoustics*, Vol. I—Part A.
[8] A. Moreau, "Detection of acoustic second harmonics in solids using a heterodyne laser interferometer," J. Acoust. Soc. Am. **98**, 2745 (1995).
[9] J. K. Na and M. A. Breazeale, "Ultrasonic nonlinear properties of lead zirconat-titanate ceramics," J. Acoust. Soc. Am. **95**, 3213 (1994).

Knowledge about typical source output influences perceived auditory distance[a] (L)

John W. Philbeck[b] and Donald H. Mershon
Department of Psychology, North Carolina State University, Raleigh, North Carolina 27695

(Received 17 July 2001; accepted for publication 26 February 2002)

Vocal effort is known to influence the judged distance of speech sound sources. The present research examined whether this influence is due to long-term experience gained prior to the experiment versus short-term experience gained from exposure to speech stimuli earlier in the same experiment. Speech recordings were presented to 192 blindfolded listeners at three levels of vocal output. Even upon the first presentation, shouting voices were reported as appearing farthest, whispered voices closest. This suggests that auditory distance perception can be affected by past experience in a way that does not require explicit comparisons between individual stimuli. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1471899]

PACS numbers: 43.66.Qp, 43.66.Lj, 43.71.Bp [LRB]

I. INTRODUCTION

Egocentric distance is the distance between an observer and a point in space; perceived egocentric distance in the auditory domain is the apparent distance between a listener and a sound source. Stimulus information that influences this perception includes the intensity of the sound reaching a listener's ears and the ratio of direct to reflected sound in a given environment (Bronkhorst and Houtgast, 1999; Mershon and King, 1975; Zahorik, 1998). In addition to these stimulus variables, a listener might also determine the source distance of familiar sounds by comparing the sound pressure level at the ears with some internal estimate of the probable output power of the sound source. For example, given a very faint proximal stimulus that one identifies as a fire engine siren, one might perceive the source to be far away, because sirens usually have high output power. Familiarity with a sound source can encompass many different kinds of information, but here we will define ''source familiarity'' more specifically to mean the stored knowledge upon which one might base such estimates of output power. As yet, only a few studies have systematically investigated the influence of sound source familiarity on the perception of auditory distance (Brungart and Scott, 2001; Gardner, 1969). This research has focused on speech sounds, and the results clearly show that when other stimulus factors are held approximately constant, estimates of the source distance of speech sounds are modulated by the production level used in generating the speech;[1] specifically, listeners indicate the source distance of whispered speech to be nearer than that of shouted speech.

A critical question remains unanswered by the foregoing research, however. The previous studies tested a relatively small number of listeners and averaged across multiple judgments per condition per listener when analyzing the data. This being the case, one cannot determine whether the source familiarity effects are due to long-term experience with the typical production level of speech or are instead due to repeated exposures to particular speech sounds within the immediate experimental setting. If the latter is true, the perceived source distance of the *first* speech sample in an experiment could be determined primarily by reverberation, absolute intensity, or some other kind of distance information that does not depend upon source familiarity (Mershon and King, 1975). Listeners might then base their distance estimates in subsequent presentations on *changes* in the perceived production level of the speech samples, relative to their initial estimate. In this way, previous reports of source familiarity effects (Brungart and Scott, 2001; Gardner, 1969) may be due to comparisons between stimuli within the immediate experimental context and have nothing to do with long-term experience with speech sounds. This letter describes a control experiment designed to rule out this possibility. To prevent individuals from comparing stimuli across trials, one must analyze the data obtained from the very first stimulus presentation. If *long-term* knowledge about the typical source output of speech contributes to the perceptual localization of the sound source relative to the listener, production level should affect source distance judgments even upon the first stimulus presentation.

II. METHOD

A. Testing environment

Testing was conducted in a carpeted room, $7.3\times7.3\times3.7$ m, with an average reverberation time (T_{60}, the time required for a sound to decay by 60 dB) of approximately 0.3 s across the frequencies of interest. The stimuli were presented by a Polk Audio (Model 5) loudspeaker system 2.5 m from the listener's head, positioned approximately at ear level in the median plane. The listener stood in front of a sound-absorbing wedge which reduced reflections from the wall behind the listener. The straight line between the listener

[a]Portions of this work were presented at the 32nd annual meeting of the Psychonomic Society, 22–24 November 1991, San Francisco, CA.

[b]Current address: Department of Psychology, The George Washington University, 2125 G Street, NW, Washington, DC 20052. Electronic mail: philbeck@gwu.edu

and loudspeaker was parallel to two walls, but slightly offset from the room's center-line. Thirty-six overhead loudspeakers (12.7 cm diameter) created a diffuse wideband masking noise to hide noise intrusions from outside the laboratory. The sound level of this noise at the listener's ears was 48 dBA. Previous work involving this room in a similar configuration demonstrated that sufficient reverberation remains to generate some modulation in distance estimates for stimuli consisting of white noise bursts (Mershon *et al.*, 1989). Median distance estimates increased by just over a meter for a range of source distances between 0.75 and 6 m; at 3 m (nearly the same source distance as in the present study), the median response was approximately 1 m. Because the source distance did not vary in the current experiment, reverberation information signaled the same source distance for each production level. Under these conditions, distance judgments could be biased toward the source distance given by reverberation (perhaps near 1 m, based on the findings of Mershon *et al.*, 1989).

B. Generation and presentation of stimuli

The sound stimuli were recorded in the testing room described above. One male and one female talker were recorded speaking the phrase "How far away from you does my voice seem?" Each talker provided a sample of the phrase using a whisper, a shout, and a normal conversational level[2] with the microphone positioned approximately 30 cm in front of the talker's mouth. For whispered recordings, the talkers whispered as if communicating with someone at the distance of the microphone; for conversational recordings, they used a voice appropriate for communicating with someone just beyond arm's reach; for shouted recordings, they attempted to shout as loudly as possible. Speech samples were digitized at 44.1 kHz with 16-bit resolution; during playback, the samples were amplified by a Crown DL-2 preamplifier and Crown PS-200 amplifier before being sent to the loudspeaker.

A Rion NA-61 Impulse Precision Sound Level Meter (with an NA-2X third-octave filter set) was used to obtain average and peak sound levels. Spectrographic analysis showed that, not surprisingly, the male voice included lower frequency components than did the female voice. The different production levels also showed clear variation. The whispered stimuli generally lacked the very low frequency energy associated with voicing and was dominated by energy in the middle and upper frequency ranges. The shouts tended to be dominated by lower frequency components associated with voicing of vowels. Conversational speech fell somewhere between these extremes.

Ideally, one would want the sound level of all stimuli to be equal in order to eliminate this as a controlling factor. We adopted the conservative approach of setting both average and peak levels of the shouting voice to be slightly *higher* than the corresponding values for the whispered and conversational voices (see Table I). This ensured that, whatever the contribution of sound level, it should have worked *against* the expected perception of a distant shout.

TABLE I. Average and peak sound levels of the three different speech samples by the male and female talkers, as presented on playback to listeners and measured at the position of the listener's head. All values given in sound levels (dBA). These stimuli were heard against a background of wide band masking noise presented at 46 dBA.

	Male voice		Female voice	
	Average	Peak	Average	Peak
Shouted	72	78	72	82
Conversational	66	74	67	77
Whispered	66	74	67	76

C. Listeners

A total of 192 listeners (96 men, 96 women) participated in this experiment for course credit. All reported normal hearing in both ears. None had previously seen the laboratory.

D. Design

Listeners were randomly divided into six groups of 32 listeners. The groups were distinguished by which talker and which production level was heard on the first presentation. Following the initial presentation of one of the six possible stimuli, each listener was then presented with the other two production levels, using the same voice (male or female) heard on the first presentation. Finally, the listener was presented again with the sample heard initially. Thus, each listener separately contributed an initial report for one of the samples, followed by additional reports for all three samples spoken in the same (male or female) voice. This design allowed for nonoverlapping analyses for first presentations and for presentations with an explicit preceding comparison stimulus. Each of the possible orderings of stimuli occurred equally often.

E. Procedure

Listeners were blindfolded before entering the testing room. They were never given prior exposure to speech from a distance within the testing environment, nor did they hear the stimulus voices before the actual stimulus presentation. The four stimuli were presented in sequence, with the listener verbally judging the source distance after each stimulus. The instructions strongly emphasized that reports should be based on the *apparent* source distance, as opposed to trying to estimate the objectively accurate distance. This methodological detail is known to enhance the influence of perceptual factors in determining the response over explicitly cognitive factors (for a review, see Carlson, 1977). The listeners presumably knew the testing environment was indoors, but the instructions were carefully worded to avoid suggesting any particular real or imagined size of the testing room. Neither vision of the workspace nor error feedback was provided until after the final response.

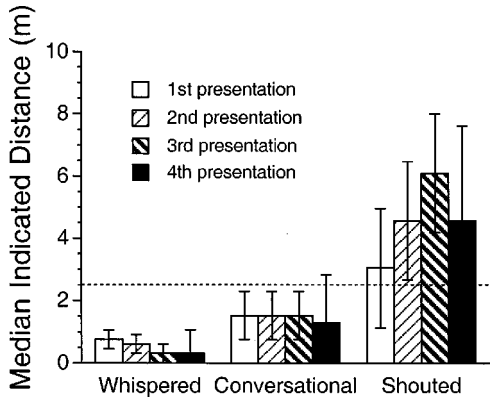


FIG. 1. Median indicated distances, showing data from presentations 1–4 for three production levels. Error bars represent ±1 semi-interquartile range. The horizontal dashed line indicates the physical distance of the loudspeaker. One estimate of 402 m (first presentation, shouted stimulus) was omitted for this figure, but otherwise each bar represents $n=64$.

F. Results

1. First presentation data

The median distance judgments for the four consecutive stimulus presentations are shown in Fig. 1. There is a clear increase in indicated distance across the three styles of speech even upon the first stimulus presentation. Median response values for whispered, conversational, and shouted speech were 0.76, 1.52, and 3.05 m, respectively, when these stimuli occurred first in the running order. The physical source distance was 2.5 m in all cases; the general inaccuracy of the responses is very likely a consequence of the limited availability of stimulus information specifying the source distance (see Mershon *et al.*, 1989). Of greater interest than the general pattern of inaccuracy, however, is the pronounced and systematic changes in distance judgments across the three production levels. Some skewing was apparent in the data, so an analysis of variance (ANOVA) using the ranks of the indicated distance values was performed. The rank transformation discards the original estimates and retains only the ordinal relations among them; the result is a test that, by making fewer assumptions about the data, is less sensitive (i.e., is *less* likely to detect differences between groups than before the transformation). This conservative analysis showed that the production level variable was significant ($F_{2,180}=59.17$, $p<0.0001$), with no other significant main effects or interactions.

2. Later presentations

When the listeners had the opportunity to make comparisons across the different stimulus presentations, the reports of distance for the whispered and shouted voices became more different. An ANOVA was performed on the ranked values of these repeated-measures data (presentations 2–4). In addition to a main effect of production level ($F_{2,360}=790.69$, $p<0.0001$), there were significant main effects of the sex of the listener ($F_{1,180}=11.17$, $p=0.001$) and of the sex of the voice used as a stimulus ($F_{1,180}=4.22$, $p=0.0415$). There was also a significant interaction of production level and the presentation order, i.e., which production level was presented on the initial trial ($F_{4,360}=4.71$, $p=0.001$). Differences between reports from male and female listeners may represent a genuinely perceptual difference or, more probably, some difference in how each group assigned numbers to a common perceptual experience. The difference associated with the sex of the talker may be related to differences in the typical output power of male shouts relative to female shouts. At present, the effects of individual differences on perceived auditory distance are poorly understood (although see Brungart and Scott, 2001, for one analysis).

III. CONCLUSIONS

There are two main conclusions. First, listeners clearly report whispers, conversational speech, and shouts at systematically different distances, even upon initial presentations and under conditions in which prior conceptions about the possible source locations are minimized. Analysis of the first presentation data of nearly 200 listeners firmly establishes that the effects of source familiarity are the result of long-term experience with speech sounds, rather than comparisons between speech stimuli encountered within the immediate experimental context. Variations in production level from shouting to whispering were associated with changes in distance judgments by as much as a factor of 4 (medians: 3.05 m vs 0.76 m). By demonstrating that source familiarity affects egocentric distance estimates even when comparisons with other experimental stimuli have been prevented, we have shown that source familiarity provides absolute distance information (Gogel, 1968; Mershon and Bowers, 1979). Second, the results shed light on the time course of the accretion of information across multiple stimulus presentations. Specifically, the median distance estimates changed systematically over a very short time scale, on the order of only a few trials. Presumably, after the listeners responded to the initial stimulus presentation, their distance estimates were influenced by a combination of two kinds of source familiarity: (1) long-term experience with speech obtained prior to the experiment, and (2) short-term experience with speech stimuli presented earlier in the experiment. Although we did not attempt to determine the relative contribution of these two sources in the trials following the initial stimulus presentation, it is clear that the effect of source familiarity was heightened when the two kinds of familiarity were available in combination (Fig. 1). The very rapid change in distance estimates upon repeated exposure to a single voice may explain why previous work has found virtually no effect for prior exposure to a talker's voice (Brungart and Scott, 2001); if such changes become attenuated very rapidly and reach a steady state, the effect will likely become more and more diluted upon additional stimulus presentations.

In studies that use direct verbal distance estimates, it is difficult to dissociate genuine perceptual influences from more abstract cognitive influences (e.g., reasoning). Even if the verbal estimates reflect a composite of perceptual and cognitive factors, however, this composite signal behaves in a very stable and predictable manner with changes in production level. Taken together, our results and those of previous researchers (Brungart and Scott, 2001; Gardner, 1969) indicate that source familiarity is indeed a potent determinant

of perceived auditory distance, operating under a variety of conditions—across many listeners and talkers, inside and outside the laboratory, with and without vision, and using both live and prerecorded speech stimuli. These factors suggest that source familiarity can be exploited successfully to convey distance information in both real and virtual environments. The relatively large perceived distances that this information is able to generate suggests that it can contribute to the guidance of human navigation on the basis of auditory information. These results also show promise for applications concerning the design of auditory displays to minimize attentional demands in high-workload situations.

ACKNOWLEDGMENTS

The authors thank William Franklin for use of the speech spectrograph, Elliott Inman and Mary Catherine Bunn for their voice work, Pat Cox for technical assistance, and David Clarke, Robert Remez, and Pavel Zahorik for helpful comments on earlier drafts.

[1]''Production level'' denotes the sound pressure level of a talker's voice measured from a fixed distance near his or her head (Brungart and Scott, 2001). A related term is ''vocal effort,'' which refers to the amount of stress or force imparted to a vocal utterance (Traunmüller and Eriksson, 2000). Stimulus information signaling vocal effort may be used to estimate the production level of a particular utterance. Vocal effort and production level are tightly linked, but in certain portions of the vocal effort continuum, the correlation is not perfect. An unvoiced ''stage whisper,'' for example, might result in a higher production level than a quietly voiced utterance, even though the whisper might entail less apparent vocal effort than the voiced speech. The linkage between vocal effort and production level is sufficiently close, however, that production level provides a useful characterization of vocal effort (Brungart and Scott, 2001).

[2]The output power of the speech samples was not directly measured, but other research (Traunmüller and Eriksson, 2000) has shown that the sound-pressure level of whispers is typically about 40 dB or less, relative to an arbitrary reference, with conversational-level speech registering at around 60 dB and shouts at 85 dB or more.

Bronkhorst, A. W., and Houtgast, T. (**1999**). ''Auditory distance perception in rooms,'' Nature (London) **397**, 517–520.

Brungart, D. S., and Scott, K. R. (**2001**). ''The effects of production and presentation level on the auditory distance perception of speech,'' J. Acoust. Soc. Am. **110**, 425–440.

Carlson, V. R. (**1977**). ''Instructions and perceptual constancy judgments,'' in *Stability and Constancy in Visual Perception: Mechanisms and Processes*, edited by W. Epstein (Wiley, New York), pp. 217–254.

Gardner, M. B. (**1969**). ''Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space,'' J. Acoust. Soc. Am. **45**, 47–53.

Gogel, W. C. (**1968**). ''The measurement of perceived size and distance,'' in *Contributions to Sensory Physiology, Vol. III*, edited by W. D. Neff (Academic, New York), pp. 125–148.

Mershon, D. H., and Bowers, J. N. (**1979**). ''Absolute and relative cues for the auditory perception of egocentric distance,'' Perception **8**, 311–322.

Mershon, D. H., and King, L. E. (**1975**). ''Intensity and reverberation as factors in the auditory perception of egocentric distance,'' Percept. Psychophys. **18**, 409–415.

Mershon, D. H., Ballenger, W. L., Little, A. D., McMurtry, P. L., and Buchanan, J. L. (**1989**). ''Effects of room reflectance and background noise on perceived auditory distance,'' Perception **18**, 403–416.

Traunmüller, H., and Eriksson, A. (**2000**). ''Acoustic effects of variation in vocal effort by men, women, and children,'' J. Acoust. Soc. Am. **107**, 3438–3451.

Zahorik, P. (**1998**). ''Experiments in Auditory Distance Perception,'' Ph.D. thesis, University of Wisconsin—Madison.

Coupled-mode sound propagation in a range-dependent, moving fluid[a)]

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

(Received 12 December 2000; revised 19 September 2001; accepted 5 February 2002)

Full-field acoustic methods for current velocity inversion require accurate and efficient mathematical models of sound propagation in a range-dependent waveguide with flow. In this paper, an exact coupled-mode representation of the acoustic field is derived. To account for the physics of the problem, normal modes in a corresponding range-independent waveguide are chosen as the local basis. In the absence of currents, mode shape functions form a complete orthogonal basis. This property is heavily used in coupled-mode theories of sound propagation in motionless fluid. Unlike in the motionless case, however, vertical dependencies of acoustic pressure in individual normal modes are not orthogonal in the presence of currents. To overcome this difficulty, linearized equations of hydrodynamics are rewritten in terms of a state vector. Its five components are expressed in terms of acoustic pressure and particle displacement due to the wave. Orthogonality of the state vectors corresponding to individual normal modes is established. Coupled differential equations are derived for range-dependent mode amplitudes, leading to a remarkably simple result. The mode-coupling equations have the same form as those known for the motionless case, but of course the values of the mode-coupling coefficients differ as long as the range dependence of the flow velocity contributes to mode coupling in addition to the range dependence of sound speed and fluid density. The mode-coupling formulation is verified against known coupled-mode equations for certain limiting cases and an exact analytic solution of a benchmark problem. [DOI: 10.1121/1.1467672]

PACS numbers: 43.20.Bi, 43.28.Py, 43.30.Bp [ANN]

I. INTRODUCTION

Environmental acoustics as well as some technological applications require the modeling of sound fields in moving media. Particularly demanding, in terms of both propagation model accuracy and computational efficiency, are full-field methods of ocean current tomography such as those discussed in Refs. 1 and 2. In response to the needs of underwater and atmospheric acoustics, most of the theoretical methods originally developed to model sound propagation in motionless media have been successfully extended to moving media (see Ref. 3 for a review). A notable exception is the modal theory of sound propagation in a range-dependent waveguide. While acoustic normal modes are known to propagate independently from one another when the waveguides are range independent (Ref. 3, Chap. 4) or their parameters change gradually (adiabatically) with range (Ref. 3, Chap. 7), no theory has been available to predict the nonadiabatic evolution of the mode spectrum due to strong, continuous range dependence in a moving medium. Coupled-mode equations that predict the mode spectrum evolution are well known and frequently used in the motionless case.[4–21] The goal of this paper is to extend the mode-coupling theory to moving media with time-independent parameters.

To satisfy the physical requirement of mode coupling vanishing in the limit of adiabatic range dependence, the normal modes of range-independent waveguides in *moving* media should be used as a local basis for representing fields in range-dependent waveguides. However, unlike in the motionless case, the depth dependence of acoustic pressure p in normal modes is not orthogonal in the presence of fluid flow. The orthogonality is heavily used in various derivations of the coupled-mode equations in the motionless acoustic waveguides (Ref. 3, Chap. 7 and Appendix B; Refs. 4–6, 9, 12, 13, 20, 21). It is the nonorthogonality of mode shape functions in moving media that have precluded, so far, an extension of mode-coupling theory to waveguides with currents. To overcome this difficulty, we will characterize the acoustic field by a judiciously chosen *state vector*, that is, a set of field quantities, rather than by a single scalar dependent variable p. The idea is to choose the state vector in such a way that state vectors corresponding to two different local modes are orthogonal, while equations of linear acoustics written in terms of the state vector retain a certain canonical form that allows for partial separation of variables. The concept of state vector is by no means new. Shevchenko[6] was probably the first to utilize a state vector (without using the term) in relation to the interaction of acoustic modes. Moreover, our approach can be viewed as an application to sound in the moving fluid of the technique developed by Maupin[11] for elastic waves in solids.

[a)]Parts of this work have been previously reported at the 16th International Congress on Acoustics/135th Meeting of the ASA (Seattle, WA, June 1998) and at the 4th International Conference on Theoretical and Computational Acoustics (Trieste, Italy, May 1999).

[b)]Also affiliated with: Acoustic Wave Propagation Laboratory, P. P. Shirshov Oceanography Institute of the Russian Academy of Sciences, Moscow 117851, Russia.
[c)]Electronic mail: Oleg.Godin@noaa.gov

The mode-coupling theory to be developed in this paper is exact in the sense that we neither require the smallness of the Mach number of the currents nor impose any limitations on rate or magnitude of variations in environmental parameters. Sound speed c, flow velocity $\mathbf{u}$, and medium density ρ in the waveguide are assumed to be sufficiently smooth functions of position that are independent of time and consistent with the equations of hydrodynamics. For brevity, we neglect the coupling of discrete modes to a continuous spectrum and restrict ourselves to a two-dimensional problem of sound propagation in a waveguide with horizontal ideal (pressure-release or rigid) boundaries or without boundaries.

With this Introduction, the remainder of the paper is organized as follows. In Sec. II, we review some properties of normal modes in a range-independent waveguide in a moving medium, and introduce a state vector. Linearized equations of hydrodynamics are cast in Sec. III in terms of the state vector as a first-order differential equation with respect to the range coordinate. Coupled-mode equations are derived in Sec. IV. Properties of the mode-coupling coefficients are discussed in Sec. V. In Sec. VI, numerical solution of the coupled-mode equations is verified against an exact solution of a benchmark problem. Alternative formulations of mode-coupling theory are compared in Sec. VII. In Sec. VIII we summarize the major findings of this work.

II. NORMAL MODES IN RANGE-INDEPENDENT MOVING MEDIA

Consider continuous acoustic waves of frequency ω propagating along a horizontal coordinate x (range) in a waveguide. We assume that flow velocity $\mathbf{u}=(u_1,0,u_3)$ and environmental parameters and the acoustic field are independent of cross-range horizontal coordinate y. Environmental parameters are functions of x and depth z only. In the case of a range-independent (layered) medium, when environmental parameters do not depend on x, the vertical component of the flow velocity is identically zero,[22] and acoustic pressure in 2D normal mode is given by Ref. 3, Chap. 4

$$p_n=f_n(z)\exp(iq_nx-i\omega t) \quad n=\pm 1,\pm 2,..., \tag{1}$$

where mode shape function $f_n(z)$ and propagation constant q_n are obtained as eigenfunctions and eigenvalues of the boundary value problem defined by a 1D wave equation

$$\frac{\partial}{\partial z}\left(\frac{1}{\rho\beta^2}\frac{\partial p}{\partial z}\right)+\left(\frac{\omega^2}{\rho c^2}-\frac{q^2}{\rho\beta^2}\right)p=0, \quad \beta=1-\frac{qu_1}{\omega}, \tag{2}$$

complemented by appropriate boundary conditions on horizontal boundaries or conditions at $|z|\to\infty$. Positive and negative indices in Eq. (1) correspond to normal modes propagating toward positive and negative x. Currents break the symmetry between waves propagating to the left and waves propagating to the right, and in moving media generally $f_n(z)\neq f_{-n}(z)$ and $q_n\neq q_{-n}$. Let us normalize mode shape functions by the condition

$$\int \frac{dz}{\rho\beta_n^3}\left[f_n^2+\frac{u_1}{\omega q_n}\left(\frac{\partial f_n}{\partial z}\right)^2\right]=1, \tag{3}$$

where $\beta_n=1-u_1q_n/\omega$. Here and below, integrals over z are taken over an entire cross section $x=\text{const}$ of the waveguide.

For normal modes of different orders, the following identity holds [Ref. 3, Eq. (4.4.55)]:

$$\int \frac{dz}{\rho}\left[\left(\frac{q_m^2}{\beta_m^2}-\frac{q_n^2}{\beta_n^2}\right)f_nf_m+\left(\frac{1}{\beta_m^2}-\frac{1}{\beta_n^2}\right)\frac{\partial f_n}{\partial z}\frac{\partial f_m}{\partial z}\right] =0, \quad m\neq n. \tag{4}$$

Using the definition of β, Eqs. (3) and (4) can be written as a single equation

$$\int \frac{dz}{\rho\beta_n^2\beta_m^2}\left[(q_m\beta_n+q_n\beta_m)f_nf_m+\frac{u_1}{\omega}(\beta_n+\beta_m)\frac{\partial f_n}{\partial z}\frac{\partial f_m}{\partial z}\right] =2q_m\delta_{mn}, \tag{5}$$

where δ_{mn} is the Kronecker symbol. In the motionless case, $\beta\equiv 1$, $q_n=-q_{-n}$, and Eq. (5) reduces to the familiar orthogonality relation of normal modes

$$\int \frac{dz}{\rho}f_nf_m=\delta_{n,m}+\delta_{n,-m}. \tag{6}$$

We have the sum of two Kronecker symbols, rather than a single Kronecker symbol, on the right-hand side of Eq. (6) because in our nomenclature, normal modes propagating to the right and to the left have mode orders of opposite signs. The identity Eq. (5) is a direct extension of Eq. (6) to moving media. It cannot, however, be used to derive mode-coupling equations in the same way that Eq. (6) is used in the motionless case (see, e.g., Ref. 3, Sec. 7.1.1). There are two reasons for this. First, the integrand in Eq. (5) contains derivatives $\partial f_n/\partial z$ of the shape functions in addition to the shape functions themselves and, second, the coefficient in front of the product f_nf_m depends on the mode orders n, m.

To cast Eq. (5) in a form more similar to Eq. (6), we need to introduce several additional concepts. Consider displacement of fluid particles in the acoustic wave. Oscillatory displacement of fluid particles[23] $\mathbf{w}$ is related to their oscillatory velocity $\mathbf{v}$ by

$$\mathbf{v}=d\mathbf{w}/dt-(\mathbf{w}\cdot\nabla)\mathbf{u}, \quad d/dt=\partial/\partial t+\mathbf{u}\cdot\nabla, \tag{7}$$

where d/dt is a convective time derivative. In a normal mode, oscillatory displacement is proportional to the acoustic pressure gradient

$$\mathbf{w}=(\rho\omega^2\beta_n^2)^{-1}\nabla p. \tag{8}$$

Further, we use $\mathbf{w}$ to define the state vector $\mathbf{\Phi}$ by the equation

$$\mathbf{\Phi}=(p,w_1,dw_1/dt,w_3,dw_3/dt)^T, \tag{9}$$

with superscript T denoting matrix transposition. For an acoustic field that is due to a single normal mode, from Eqs. (1), (8), and (9) we have $\Phi=\Phi_m(z)\exp(iq_mx-i\omega t)$

$$\mathbf{\Phi}_m=\left(f_m,\frac{iq_mf_m}{\rho\omega^2\beta_m^2},\frac{q_mf_m}{\rho\omega\beta_m},\frac{1}{\rho\omega^2\beta_m^2}\frac{\partial f_m}{\partial z},\frac{-i}{\rho\omega\beta_m}\frac{\partial f_m}{\partial z}\right)^T. \tag{10}$$

Consider the acoustic field in a waveguide with *reversed flow*; that is, a waveguide with a flow velocity profile $\tilde{\mathbf{u}}(z)=-\mathbf{u}(z)$ and with the same boundaries as well as sound-

speed and density profiles $c(z)$ and $\rho(z)$ as in the original waveguide.[24] Modes of the original waveguide and the one with reversed flow are connected by the obvious relations $\tilde{q}_{-n}(\omega)=-q_n(\omega)$ and $\tilde{f}_{-n}(z;\omega)=f_n(z;\omega)$, $n=\pm1,\pm2,\ldots$. Therefore, for the state vector $\mathbf{\Psi}_n$ of a mode in a waveguide with reversed flow, we have

$$\mathbf{\Psi}_{-n}=\left(f_n,\frac{-iq_nf_n}{\rho\omega^2\beta_n^2},\frac{-q_nf_n}{\rho\omega\beta_n},\frac{1}{\rho\omega^2\beta_n^2}\frac{\partial f_n}{\partial z},\frac{-i}{\rho\omega\beta_n}\frac{\partial f_n}{\partial z}\right)^T. \tag{11}$$

The mode state vector Eq. (11) in a medium with reversed flow can be obtained by simultaneously changing the signs of mode order m and current velocity $\mathbf{u}$ in the mode state vector Eq. (10) in the original medium.

In terms of mode state vectors $\mathbf{\Phi}_m$ Eq. (10) and $\mathbf{\Psi}_{-n}$ Eq. (11), Eq. (5) expressing normal-mode orthogonality in moving media becomes

$$\int dz(\mathbf{\Psi}_{-\mathbf{n}})^T[B]\Phi_{\mathbf{m}}=-2i\omega^{-2}q_m\delta_{mn}, \tag{12}$$

where $[B]$ is a 5×5 matrix with just six nonzero elements

$$[B]=\begin{pmatrix}0&-1&0&0&0\\1&0&\rho u_1&0&0\\0&\rho u_1&0&0&0\\0&0&0&0&\rho u_1\\0&0&0&\rho u_1&0\end{pmatrix}. \tag{13}$$

The mode orthogonality relation Eq. (12) is largely similar to its counterpart Eq. (6) in the quiescent medium but differs in two respects. Modes are characterized in Eq. (12) by vectors in the moving medium rather than scalar shape functions in Eq. (6); the mode state vectors are orthogonal with a matrix weight $[B]$ rather than scalar weight $1/\rho$ in quiescent media. In both cases, the weights depend on z but not on mode indices.

III. EQUATIONS OF MOTION

Our goal in this section is to cast equations of linear acoustics in range-dependent, moving fluid as a vector differential equation of the first order with respect to coordinate x. The starting point of our analysis is equations of motion in terms of acoustic pressure and oscillatory displacement[23]

$$\rho d^2\mathbf{w}/dt^2+\nabla p+(\mathbf{w}\cdot\nabla)\nabla p_0-(\rho c^2)^{-1}\times\nabla p_0(p+\mathbf{w}\cdot\nabla p_0)=0, \tag{14}$$

$$\nabla\cdot\mathbf{w}+(p+\mathbf{w}\cdot\nabla p_0)/\rho c^2=0. \tag{15}$$

Equations (14) and (15) have been obtained by linearizing equations of hydrodynamics with respect to wave amplitude and describe waves in the rather general case of an arbitrary inhomogeneous, multicomponent, moving fluid. Current velocity enters Eq. (14) through a convective time derivative. Sound speed, flow velocity, medium density, and pressure p_0 in the absence of the wave are related by equations[23]

$$\nabla p_0=-\rho(\mathbf{u}\cdot\nabla)\mathbf{u}, \tag{16a}$$

$$\nabla\cdot(\rho\mathbf{u})=0, \tag{16b}$$

$$(\mathbf{u}\cdot\nabla)p_0=c^2(\mathbf{u}\cdot\nabla)\rho. \tag{16c}$$

We ignore the effects of gravity.[26] In layered (i.e., range-independent) media, where $u_3\equiv0$ and $\mathbf{u}=\mathbf{u}(z)$, unperturbed pressure $p_0=\text{const}$ according to Eq. (16a) and Eq. (8) follows immediately from Eq. (14).

In the general case, from Eq. (15) and the definition of the convective time derivative, we have

$$\partial w_1/\partial x=-\partial w_3/\partial z-(\rho c^2)^{-1}(p+\mathbf{w}\cdot\nabla p_0), \tag{17}$$

$$\frac{\partial w_3}{\partial x}=\frac{1}{u_1}\left(i\omega w_3-u_3\frac{\partial w_3}{\partial z}+\frac{dw_3}{dt}\right). \tag{18}$$

From the z component of Eq. (14), it follows that

$$\frac{\partial}{\partial x}\left(\frac{dw_3}{dt}\right)=\frac{1}{u_1}\left[i\omega\frac{dw_3}{dt}-u_3\frac{\partial}{\partial z}\left(\frac{dw_3}{dt}\right)-\frac{1}{\rho}\frac{\partial p}{\partial z}-\frac{1}{\rho}(\mathbf{w}\cdot\nabla)\frac{\partial p_0}{\partial z}+\frac{p+\mathbf{w}\cdot\nabla p_0}{\rho^2c^2}\frac{\partial p_0}{\partial z}\right]. \tag{19}$$

Applying the operator d/dt to Eq. (17) and using the x component of Eq. (14), we find after using some algebra that

$$\frac{\partial}{\partial x}\frac{dw_1}{dt}=(1-u_1^2/c^2)^{-1}\left(Q_1-\frac{u_1}{c^2}Q_2\right), \tag{20}$$

$$\frac{\partial p}{\partial x}=\rho(1-u_1^2/c^2)^{-1}(Q_2-u_1Q_1), \tag{21}$$

where

$$\begin{aligned}Q_1=&\left(i\omega-\frac{\partial u_1}{\partial x}-u_3\frac{\partial}{\partial z}\right)\left(\frac{\partial w_3}{\partial z}+\frac{p+\mathbf{w}\cdot\nabla p_0}{\rho c^2}\right)+\frac{\partial u_3}{\partial x}\frac{\partial w_1}{\partial z}+\left(\frac{1}{u_1}\frac{\partial u_1}{\partial z}-\frac{1}{\rho c^2}\frac{\partial p_0}{\partial x}-\frac{\partial}{\partial z}\right)\left(i\omega w_3\right.\\&\left.-u_3\frac{\partial w_3}{\partial z}+\frac{dw_3}{dt}\right)-u_1\left[p\frac{\partial}{\partial x}\frac{1}{\rho c^2}-\frac{1}{\rho c^2}\frac{\partial p_0}{\partial x}\frac{\partial w_3}{\partial z}+w_1\frac{\partial}{\partial x}\left(\frac{1}{\rho c^2}\frac{\partial p_0}{\partial x}\right)+w_3\frac{\partial}{\partial x}\left(\frac{1}{\rho c^2}\frac{\partial p_0}{\partial z}\right)\right.\\&\left.-\frac{1}{\rho^2c^4}\frac{\partial p_0}{\partial x}(p+\mathbf{w}\cdot\nabla p_0)\right],\end{aligned} \tag{22}$$

$$Q_2=i\omega\frac{dw_1}{dt}-u_3\frac{\partial}{\partial z}\frac{dw_1}{dt}-\frac{1}{\rho}(\mathbf{w}\cdot\nabla)\frac{\partial p_0}{\partial z}+\frac{p+\mathbf{w}\cdot\nabla p_0}{\rho^2c^2}\frac{\partial p_0}{\partial z}. \tag{23}$$

In the transformations, the relations Eq. (16) between environmental parameters have been used.

Equations (17)–(23) give range derivatives of the five components of the state vector $\mathbf{\Phi}$ in terms of environmental parameters and values of the $\mathbf{\Phi}$ components at a given cross section of the waveguide. In a vector form, Eqs. (17) to (23) become

$$\partial\mathbf{\Phi}/\partial x=[A]\mathbf{\Phi}, \tag{24}$$

where $[A]$ is a 5×5 matrix that contains parameters of the

medium (and their derivatives) as well as operators $\partial/\partial z$ and $\partial^2/\partial z^2$, but does not contain operators $\partial/\partial x$ acting on $\mathbf{\Phi}$. The matrix $[A]$ does not depend on $\mathbf{\Phi}$; that is, Eq. (24) is linear. Below, we will need the matrix $[C]=[B][A]$, which admits a more compact representation than $[A]$. From Eqs. (17)–(24) and the definition Eq. (13) of $[B]$, we find

$$[C]=\begin{pmatrix} \frac{1}{\rho c^2} & \frac{1}{\rho c^2}\frac{\partial p_0}{\partial x} & 0 & \frac{1}{\rho c^2}\frac{\partial p_0}{\partial x}+\frac{\partial}{\partial z} & 0 \\ \frac{1}{\rho c^2}\frac{\partial p_0}{\partial x} & \frac{1}{\rho c^2}\left(\frac{\partial p_0}{\partial x}\right)^2-\frac{\partial^2 p_0}{\partial x^2} & i\omega\rho-\rho u_3\frac{\partial}{\partial z} & \frac{1}{\rho c^2}\frac{\partial p_0}{\partial x}\frac{\partial p_0}{\partial z}-\frac{\partial^2 p_0}{\partial x\partial z} & 0 \\ \frac{-u_1}{c^2} & \frac{-u_1}{c^2}\frac{\partial p_0}{\partial x} & 0 & \frac{-u_1}{c^2}\frac{\partial p_0}{\partial x}-\rho u_1\frac{\partial}{\partial z} & 0 \\ \frac{1}{\rho c^2}\frac{\partial p_0}{\partial x}-\frac{\partial}{\partial z} & \frac{1}{\rho c^2}\frac{\partial p_0}{\partial x}\frac{\partial p_0}{\partial z}-\frac{\partial^2 p_0}{\partial x\partial z} & 0 & \frac{1}{\rho c^2}\frac{\partial p_0}{\partial x}\frac{\partial p_0}{\partial z}-\frac{\partial^2 p_0}{\partial z^2} & i\omega\rho-\rho u_3\frac{\partial}{\partial z} \\ 0 & 0 & 0 & i\omega\rho-\rho u_3\frac{\partial}{\partial z} & \rho \end{pmatrix}. \tag{25}$$

Matrix $[C]$ proves to be much more concise than $[A]$ because, as is readily seen from Eqs. (13), (20), and (21), all terms involving Q_1 Eq. (22) cancel in $[C]$.

IV. COUPLED-MODE EQUATIONS

Consider a vertical cross section $x=x_0$ of a range-dependent waveguide. Define *local modes* as the modes of the range-independent waveguide with parameters $\rho(x_0,z)$, $c(x_0,z)$, and $\mathbf{u}=(u_1(x_0,z),0,0)$. Like ρ, c, and $\mathbf{u}$, the vector $\mathbf{\Phi}_m(x,z)$ of such a local mode is a smooth function of range. Let us represent the acoustic field $\mathbf{\Phi}$ as a sum of local modes $\mathbf{\Phi}_m(x,z)$ with as-yet-unknown mode amplitudes $F_m(x)$

$$\mathbf{\Phi}(x;z)=\sum_m F_m(x)\mathbf{\Phi}_m(z;x). \tag{26}$$

We substitute Eq. (26) into Eq. (24), multiply the resulting equation

$$\sum_m \mathbf{\Phi}_m(x,z)dF_m/dx$$

$$=\sum_m F_m(x)([A]\mathbf{\Phi}_m(x,z)-\partial\Phi_m/\partial x) \tag{27}$$

by $(\mathbf{\Psi}_{-n})^T[B]$ from the left, and integrate over z across the waveguide cross section $x=\text{const}$. Using the mode orthogonality relation Eq. (12), we obtain

$$\frac{dF_n}{dx}=\sum_m\left(\int dz(\Psi_{-n})^T[C]\mathbf{\Phi}_m - \int dz(\mathbf{\Psi}_{-n})^T[B]\frac{\partial\mathbf{\Phi}_m}{\partial x}\right)F_m(x). \tag{28}$$

Thus, we have reduced the problem to a set of coupled equations for mode amplitudes.

The form of coupled-mode equations in Eq. (28) is not well suited for theoretical analysis. It is also inconvenient for numerical implementation because evaluation of mode coupling coefficients would require knowledge of the range derivative of mode shape functions [see the second integral on the right-hand side of Eq. (28)]. Using Eqs. (13) and (25) for matrices $[B]$ and $[C]$, after tedious but simple transformations we obtain from Eqs. (28), (10), and (11)

$$\frac{dF_n}{dx}=\left(iq_n-\frac{1}{2q_n}\frac{dq_n}{dx}\right)F_n(x)+\sum_m g_{nm}(x)F_m(x), \tag{29}$$

where mode-coupling coefficients g_{nm} are given by

$$g_{nm}=\int dz(u_3E^{(1)}_{nm}+E^{(2)}_{nm}+E^{(3)}_{nm})+\frac{1}{2q_n(q_n-q_m)} \times\int dz\left[f_nf_m\frac{\partial}{\partial x}\left(\frac{\omega^2}{\rho c^2}\right)-\left(q_nq_mf_nf_m+\frac{\partial f_n}{\partial z}\frac{\partial f_m}{\partial z}\right) \times\left(\frac{1}{\beta_n^2}\frac{\partial}{\partial x}\frac{1}{\rho}+\frac{q_m(\beta_n+\beta_m)}{\omega\rho\beta_n^2\beta_m^2}\frac{\partial u_1}{\partial x}\right)\right], \quad n\neq m, \tag{30}$$

$$g_{nn}=\frac{1}{2\omega q_n}\int\frac{dz}{\rho^2\beta_n^3}\left[q_n^2f_n^2+\left(\frac{\partial f_n}{\partial z}\right)^2\right]\frac{\partial}{\partial x}(\rho u_1) + \int dz(u_3E^{(1)}_{nn}+E^{(2)}_{nn}+E^{(3)}_{nn}), \tag{31}$$

with

$$E^{(1)}_{nm}=\frac{-1}{2\omega q_n}\left[\frac{q_nq_mf_n}{\beta_n^2}\frac{\partial}{\partial z}\left(\frac{f_m}{\rho\beta_m}\right)+\frac{1}{\beta_n^2}\frac{\partial f_n}{\partial z}\frac{\partial}{\partial z}\left(\frac{1}{\rho\beta_m}\frac{\partial f_m}{\partial_z}\right) + f_m\frac{\partial f_n}{\partial z}\frac{q_m^2c^2-\omega^2\beta_m^2}{\rho c^2\beta_n\beta_m^2}\right], \tag{32}$$

$$E_{nm}^{(2)}=\frac{1}{2q_n\rho^2\omega^2\beta_n^2\beta_m^2}\left\{\left(q_nf_n\frac{\partial f_m}{\partial z}-q_mf_m\frac{\partial f_n}{\partial z}\right)\right.$$
$$\times\left(\frac{1}{\rho c^2}\frac{\partial p_0}{\partial x}\frac{\partial p_0}{\partial z}-\frac{\partial^2 p_0}{\partial x\partial z}\right)$$
$$+iq_nq_mf_nf_m\left[\frac{1}{\rho c^2}\left(\frac{\partial p_0}{\partial x}\right)^2-\frac{\partial^2 p_0}{\partial x^2}\right]$$
$$\left.+i\frac{\partial f_n}{\partial z}\frac{\partial f_m}{\partial z}\left[\frac{1}{\rho c^2}\left(\frac{\partial p_0}{\partial z}\right)^2-\frac{\partial^2 p_0}{\partial z^2}\right]\right\}, \tag{33}$$

$$E_{nm}^{(3)}=\frac{1}{2q_n\rho^2c^2\beta_n^2\beta_m^2}\left[(q_n\beta_m^2-q_m\beta_n)f_nf_m\frac{\partial p_0}{\partial x}\right.$$
$$\left.+i\left(\beta_nf_n\frac{\partial f_m}{\partial z}+\beta_m^2f_m\frac{\partial f_n}{\partial z}\right)\frac{\partial p_0}{\partial z}\right]. \tag{34}$$

The mode orthogonality relation Eq. (5) as well as the relation (Ref. 3, Sec. 4.5.1)

$$\frac{dq_n}{dx}=\frac{1}{2q_n}\int\frac{dz}{\rho\beta_n^3}\left\{\left[\left(q_n^2-\frac{\omega^2\beta_n^2}{c^2}\right)f_n^2+\left(\frac{\partial f_n}{\partial z}\right)^2\right]\frac{\beta_n}{\rho}\frac{\partial\rho}{\partial x}\right.$$
$$\left.-\frac{2\omega^2\beta_n{}^3}{c^3}f_n^2\frac{\partial c}{\partial x}-\frac{2q_n}{\omega}\left[q_n^2f_n^2+\left(\frac{\partial f_n}{\partial z}\right)^2\right]\frac{\partial u_1}{\partial x}\right\} \tag{35}$$

between environmental gradients and variations of local mode propagation constant have been utilized in deriving the mode-coupling equations (29).

In the range-independent waveguide, $dq_n/dx=0$ and $g_{nm}=0$. From the coupled-mode equations (29) we have, then, $F_n(x)=\exp[iq_n(x-x_0)]F_n(x_0)$; that is, each mode propagates independently with its amplitude determined solely by initial conditions and its propagation constant q_n, as expected. In Eqs. (29) coefficients g_{nm} describe coupling between various modes and are generally nonzero in range-dependent waveguides. The second integral in Eq. (30) depicts explicit contributions to mode coupling due to range variation of sound speed, current velocity, and medium density. In moving media, additional contributions to mode coupling are due to the vertical component u_3 of the flow velocity and variations in the unperturbed pressure p_0; see Eqs. (30) and (32)–(34). According to Eq. (16b), the vertical velocity is nonlocally related to range derivatives of density and horizontal velocity

$$u_3(x,z)=\frac{-1}{\rho(x,z)}\int_H^z dz'\frac{\partial}{\partial x}[\rho(x,z')u_1(x,z')], \tag{36}$$

where the lower limit of integration H can be taken at a waveguide boundary or at $-\infty$ for unbounded waveguides. Like u_3, the gradient of the unperturbed pressure is another indirect manifestation of the medium range dependence. According to Eq. (16a), ∇p_0 vanishes when u_3 and $\partial u_1/\partial x$ vanish.

The diagonal elements g_{nn} Eq. (31) of the mode-coupling matrix can be viewed as a correction, due to the range dependence, to the real and imaginary parts of the mode propagation constant q_n.

V. PROPERTIES OF MODE-COUPLING COEFFICIENTS

Let us now discuss some properties of the mode-coupling coefficients and analyze the general coupled-mode equations (29) for several limiting cases. The cases are chosen to illustrate the relation of the coupled-mode formulation for moving media, derived above, to previous results.

A. The case of vanishing flow

Consider sound propagation in a waveguide with $\mathbf{u}\to 0$. According to Eq. (16a), $\nabla p_0\to 0$ in this case, and the mode-coupling coefficients Eqs. (30) and (31) become

$$g_{nn}=0,$$
$$g_{nm}=\frac{1}{2q_n(q_n-q_m)}\int dz\left[f_nf_m\frac{\partial}{\partial x}\left(\frac{\omega^2}{\rho c^2}\right)\right.$$
$$\left.-\left(q_nq_mf_nf_m+\frac{\partial f_n}{\partial z}\frac{\partial f_m}{\partial z}\right)\frac{\partial}{\partial x}\frac{1}{\rho}\right],\quad n\neq m. \tag{37}$$

Let us compare this result to known coupling equations in motionless fluid. There are two kinds of acoustic coupled-mode equations for waveguides with continuously varying parameters. In the first approach, no distinction is made between modes propagating in opposite directions. This approach[4,5] relies on mode shape functions being independent of propagation direction in motionless fluid, and leads to second-order equations for mode amplitudes. In the second approach,[6,9] amplitudes of modes propagating in opposite directions are considered individually, leading to first-order coupled equations for these amplitudes. The two representations are equivalent, and one can easily be derived from the other (see Ref. 3, Sec. 7.1 and Appendix B). Obviously, the second representation is better suited for a comparison with coupled-mode equations in moving fluid. Indeed, in waveguides in moving media, modes propagating in opposite directions generally have distinct shape functions and, therefore, have to be individually assigned range-dependent mode amplitudes.

Using the notational conventions of this paper, the first-order mode-coupling equations for a motionless medium, as given by Eqs. (7.1.33) and (7.1.34) of Ref. 3, transform into Eq. (29) with the coupling coefficients

$$g_{nm}=\frac{q_m}{2q_n}D_{mn}-\frac{1}{2}D_{nm},\quad D_{nm}=\int\frac{dz}{\rho}f_n\frac{\partial f_m}{\partial x}. \tag{38}$$

It follows immediately from Eq. (38) that $g_{nn}=0$, in agreement with Eq. (37). For a waveguide without boundaries or with horizontal ideal boundaries, D_{nm} is given in terms of environmental gradients by Eq. (7.1.17) of Ref. 3

$$D_{nm}=\frac{1}{q_m^2-q_n^2}\int\frac{dz}{\rho}\left[\left(\omega^2\frac{\partial c^{-2}}{\partial x}-\frac{\omega^2-q_m^2c^2}{\rho c^2}\frac{\partial\rho}{\partial x}\right)f_nf_m\right.$$
$$\left.+\frac{1}{\rho}\frac{\partial\rho}{\partial x}\frac{\partial f_n}{\partial z}\frac{\partial f_m}{\partial z}\right],\quad n\neq m. \tag{39}$$

Inspection reveals that when $n\neq m$, the mode-coupling coefficients defined by Eqs. (38) and (39) are identical to the mode-coupling coefficients defined by Eq. (37).

Hence, as expected, the mode-coupling equations (29) reduce to the known mode-coupling equations for motionless media in the limit of vanishing flow.

B. Adiabatic approximation

When parameters of the waveguide vary gradually with range, according to Eq. (36), the vertical component of the flow velocity u_3 is inversely proportional to the spatial scale L of the range dependence, while $\partial p_0/\partial x \sim 1/L$ and $\partial p_0/\partial z \sim 1/L^2$ according to Eq. (16a). With derivatives $\partial u_1/\partial x$, $\partial c/\partial x$, and $\partial \rho/\partial x$ all proportional to $1/L$, the coupling coefficients g_{nm}, as defined by Eqs. (30)–(34), become negligible in the limit of large L. Then, equations for mode amplitudes of different orders decouple in Eq. (29), and we obtain

$$F_n(x) = \sqrt{\frac{q_n(x_0)}{q_n(x)}} \exp\left(i \int_{x_0}^{x} q_n(x')dx' \right) F_n(x_0), \qquad (40)$$

where x_0 is an arbitrarily chosen, initial cross section of the waveguide. This result coincides with the well-known *adiabatic approximation* for sound in a moving medium with gradual range dependence, previously derived from other considerations (Ref. 3, Sec. 7.3). For two-dimensional problems, it has the same form as in the motionless case; compare Eqs. (40) and (7.1.40) in Ref. 3. The factor $[q_n(x_0)/q_n(x)]^{1/2}$ in front of the exponent on the right side of Eq. (40) originates from the term $(-\frac{1}{2} q_n{}^{-1} dq_n/dx)F_n$ in Eq. (29). Although the term is small for large L, keeping it in the equation for mode amplitude is essential. It can be shown that the factor $[q_n(x_0)/q_n(x)]^{1/2}$ in Eq. (40) ensures acoustic energy conservation within the adiabatic approximation.

When the coefficients g_{nm} in Eq. (29) are small but not negligible—that is, mode coupling is weak—solution of the coupled-mode equations can be found (and, further, quantitative conditions of validity of the adiabatic approximations can be obtained) by the method of successive approximations. For the motionless case, such an analysis is presented in Ref. 3, Sec. 7.1.3. With the coupled-mode equations in moving and motionless media being the same except for values of the coupling coefficients, the analysis can be extended to the moving media in a straightforward manner and need not be reproduced here.

C. Symmetries of the coupling coefficients

The physical process of mode coupling cannot be affected by our choice of the reference frame in which this process is considered. The invariance with respect to the choice of the reference frame imposes some *a priori* restrictions on the coupling coefficients which can be used to verify Eqs. (30)–(34).

1. Range coordinate reversal

If the direction of the Ox coordinate axis is reversed, environmental parameters p_0, c, ρ, and u_3 preserve their values, while u_1 becomes $-u_1$. The mode having mode order n in the original reference frame will be characterized by the mode order $-n$ because it now propagates in the opposite direction. Characteristics of the local modes transform according to $q_n \to -q_{-n}$, $f_n \to f_{-n}$, and $\beta_n \to \beta_{-n}$. Hence, for the coupled-mode equations (29) to predict the same acoustic field in the two reference frames, the coupling coefficients should transform as $g_{n,m} \to -g_{-n,-m}$ for all n and m.

Given the above transformation rules for the environmental and local-mode parameters and that $\partial/\partial x \to -\partial/\partial x$, it is easy to verify that the coefficients $E_{nm}^{(j)}$, $j=1,2,3$ defined by Eqs. (32)–(34) transform as $E_{nm}^{(j)} \to -E_{-n,-m}^{(j)}$. An inspection of Eqs. (30) and (31) then shows that coupling coefficients $g_{n,m}$ indeed become $-g_{-n,-m}$ upon the range coordinate reversal, as required by the acoustic field invariance.

2. Depth coordinate reversal

If the direction of the Oz coordinate axis is reversed, mode orders, shape functions, and propagation constants remain unchanged as well as the environmental parameters, except for u_3; $u_3 \to -u_3$, $\partial/\partial z \to -\partial/\partial z$. For the acoustic field obtained by solving the coupled-mode equations (29) to be invariant with respect to the reference frame transformation, the coupling coefficients g_{nm} should not change.

Note that $E_{nm}^{(1)}$ [Eq. (32)] contains only odd powers of $\partial/\partial z$, while $E_{nm}^{(2)}$ [Eq. (33)] and $E_{nm}^{(3)}$ [Eq. (34)] contain only even powers of the differential operator; u_3 does not enter Eqs. (32)–(34). An inspection then shows that both diagonal Eq. (31) and off-diagonal Eq. (30) mode-coupling coefficients are indeed invariant with respect to the depth coordinate reversal, as required.

3. Time reversal

Changing the sign of time corresponds to the following transformation of flow velocity and wave frequency: $\mathbf{u} \to -\mathbf{u}$, $\omega \to -\omega$. All the modes change their direction of propagation but not their phase and amplitude dependence on position. Hence, $q_n \to q_{-n}$, $f_n \to f_{-n}$, $\beta_n \to \beta_{-n}$, and $F_n \to F_{-n}$. For the coupled-mode equations (29) to predict the same acoustic field, the coupling coefficients should transform as $g_{n,m} \to g_{-n,,-m}$.

To establish compliance of the coupling coefficients Eqs. (30) and (31) with this requirement, it is sufficient to show that the coupling coefficients are invariant with respect to simultaneous change in signs of $\mathbf{u}$ and ω, with all other quantities in Eqs. (30)–(34) unchanged. Note that $E_{nm}^{(2)}$ in Eq. (33) and $E_{nm}^{(3)}$ in Eq. (34) are proportional to even powers of ω, while $E_{nm}^{(1)}$ in Eq. (32) is proportional to $1/\omega$; $E_{nm}^{(j)}$ do not explicitly depend on $\mathbf{u}$. It follows then from Eqs. (30) and (31) that g_{nm} is invariant with respect to a simultaneous change in the sign of $\mathbf{u}$ and ω. Hence, the coupling coefficients calculated in Sec. IV have the required symmetry.

D. Effective quiescent medium

Under certain conditions, sound propagation in a medium where the ratio of the current velocity to sound speed $M=u/c$ is small can be modeled as propagation in an "effective" motionless fluid. Such an approach is usually loosely referred to as an *effective sound-speed approximation*. In fact, a full account of first-order terms in M requires introduction of an effective quiescent fluid where density as

well as sound speed depends on the velocity of current in the original medium.[27] Therefore, we will use the term *effective motionless medium approximation (EMMA)*. In the literature, applicability conditions of various versions of EMMA have been addressed mostly for stratified media.[28–30] Theoretical justification of EMMA and analyses of its conditions of validity in range-dependent environments have been put forward only in the contexts of either the ray theory[27] or the parabolic approximation.[31,32] The validity of EMMA can be established without invoking the ray and parabolic approximations by directly comparing acoustic wave equations for moving and quiescent media. It is shown in Ref. 33 that, when environmental parameters vary weakly or gradually in the direction **N** of sound propagation, and the direction itself is little changed over distances of the order of wavelength, the wave equation for sound in a moving medium with $M \ll 1$ reduces to the wave equation in motionless fluid with effective sound speed and density, defined as follows:

$$c_{\text{eff}} = c + \mathbf{u}\cdot\mathbf{N}, \quad \rho_{\text{eff}} = \rho c^2/c_{\text{eff}}^2, \tag{41}$$

where **N** is a unit vector. The effective medium is independent of wave frequency but generally depends on the prevailing direction of sound propagation.

Here, we use EMMA to check the coupled-mode equations (29) in the case of weak yet nonzero currents, against known coupled-mode equations for the motionless case. Consider a range-dependent waveguide where grazing angles χ pertinent to all propagating modes are small ($\chi \ll 1$), relative variations of sound speed and density are small [$\delta c/c = O(\epsilon)$, $\delta\rho/\rho = O(\epsilon)$, $0 < \epsilon \ll 1$], and backscattering is negligible. Then, the validity conditions of EMMA are met as long as $M \ll 1$, and one can take $\mathbf{N} = \nabla x$ in Eq. (41) for waves propagating to the right. We will show that the coupled-mode equations (29) in moving fluid agree with the coupled-mode equations in the effective quiescent medium within EMMA's domain of validity.

First, let us compare mode propagation constants in the moving and effective media. Using a perturbation theory (Ref. 3, Sec. 4.5.1) with respect to a motionless medium with sound speed c and density ρ, we obtain

$$q_m - q_m^{(0)} = -\int dz\, \frac{u_1}{\rho\omega}\left[(\partial f_m^{(0)}/\partial z)^2 + (q_m^{(0)} f_m^{(0)})^2\right] + O(M^2), \tag{42}$$

$$\xi_m - q_m^{(0)} = -\int dz\, \frac{u_1}{q_m^{(0)}\rho c}\left[(\partial f_m^{(0)}/\partial z)^2 + (q_m^{(0)} f_m^{(0)})^2\right] + O(M^2). \tag{43}$$

Here, ξ_m is the propagation constant of the mth local mode in the effective medium. Superscripts 0 indicate propagation constants and mode shape functions in the unperturbed waveguide. Taking into account that $k - q_m = q_m O(\chi^2)$, we have from Eqs. (42) and (43)

$$\xi_m - q_m = q_m O(M^2 + M\chi^2) = (q_m - q_m^{(0)}) O(M + \chi^2). \tag{44}$$

Hence, EMMA correctly predicts the leading order of the effect of currents on mode propagation constants.

Mode shape functions ψ_m in the effective media Eq. (41) satisfy the one-dimensional wave equations

$$\frac{\partial}{\partial z}\left(\frac{1}{\rho_{\text{eff}}}\frac{\partial p}{\partial z}\right) + \left(\frac{\omega^2}{\rho c^2} - \frac{q_m^2}{\rho_{\text{eff}}}\right) p = 0, \tag{45}$$

and the same boundary conditions as those imposed on mode shape functions f_m in a moving medium. According to Eqs. (41) and (44), coefficients of Eq. (45) differ from coefficients of Eq. (2) for f_m by small terms $O(M^2 + M\chi^2)$. Taking into account that mode shape functions are normalized by distinct conditions Eq. (3) and Eq. (6) in moving and motionless media, we have

$$\psi_m(z;x) = A_m(x) f_m(z;x)[1 + O(M^2 + M\chi^2)], \quad A_m = 1 + O(M). \tag{46}$$

Now we are prepared to compare mode-coupling coefficients. In the effective medium, the coupling coefficients G_{nm} are given by Eq. (37), where ξ_n, ψ_n, ρ_{eff} and c_{eff} are to be substituted for q_n, f_n, ρ and c. Using Eqs. (41), (44), and (46), for off-diagonal coefficients, we find

$$G_{nm} = \frac{1 + O(M)}{2\xi_n(\xi_n - \xi_m)} \int dz \left[f_n f_m \frac{\partial}{\partial x}\left(\frac{\omega^2}{\rho c^2}\right) - \left(q_n q_m f_n f_m + \frac{\partial f_n}{\partial z}\frac{\partial f_m}{\partial z}\right)\left(\frac{\partial}{\partial x}\frac{1}{\rho} + \frac{2}{\rho c}\frac{\partial u_1}{\partial x}\right)\right]. \tag{47}$$

Using Eqs. (16a) and (36) to evaluate ∇p_0 and u_3 in the moving medium considered, we obtain from Eqs. (30) and (32)–(34)

$$g_{nm} = \frac{1 + O(M + \epsilon + \chi)}{2q_n(q_n - q_m)} \int dz \left[f_n f_m \frac{\partial}{\partial x}\left(\frac{\omega^2}{\rho c^2}\right) - \left(q_n q_m f_n f_m + \frac{\partial f_n}{\partial z}\frac{\partial f_m}{\partial z}\right) \times \left(\frac{1}{\beta_n^2}\frac{\partial}{\partial x}\frac{1}{\rho} + \frac{q_m(\beta_n + \beta_m)}{\omega\rho\beta_n^2\beta_m^2}\frac{\partial u_1}{\partial x}\right)\right], \quad n \neq m. \tag{48}$$

Inspection of the coefficients in front of the range derivatives of sound speed, density, and horizontal current velocity in the right-hand sides of Eqs. (47) and (48) shows that contributions to mode coupling due to range variation of each of the three environmental parameters are the same, up to the factor $1 + O(M + \chi + \epsilon)$, in moving and effective media.

Finally, diagonal coefficients in the mode-coupling equations (29) in moving and effective media are $iq_n - (2q_n)^{-1} dq_n/dx + g_{nn}$ and $i\xi_n - (2\xi_n)^{-1} d\xi_n/dx$ as $G_{nn} = 0$. The terms iq_n and $i\xi_n$ in the diagonal coefficients differ by the small quantity $O(M^2 + M\chi^2)$; see Eq. (44). Comparison of the remaining terms is similar to the above comparison of the off-diagonal coefficients. Using Eq. (35) and its counterpart for the effective medium, one finds that $(2\xi_n)^{-1} d\xi_n/dx = [(2q_n)^{-1} dq_n/dx - g_{nn}][1 + O(M + \epsilon)]$. This completes verification of the coupled-mode formulation in moving media against EMMA.

VI. A BENCHMARK PROBLEM: ROTATED WAVEGUIDE

The goal of this section is to further verify results of Secs. II–IV by comparing numerical solutions of the mode-coupling equations (29) to an exact solution of a benchmark problem. We are not aware of any published exact solutions, either analytical or numerical, for range-dependent waveguides in a moving fluid with a continuous dependence of its parameters on coordinates. To generate a benchmark solution, we will use a *rotated waveguide* which is range independent in a rotated Cartesian coordinate system (x_1, z_1) but becomes range dependent in the original coordinates (x, z). The rotated coordinates are related to the original coordinates by

$$x_1 = x\cos\alpha + z\sin\alpha, \quad z_1 = -x\sin\alpha + z\cos\alpha, \tag{49}$$

where α is the angle between x and x_1 coordinate axes.

Let a waveguide be described in the rotated coordinates by a parabolic profile of the refraction index squared

$$c_0^2/c^2(z_1) = b - a^2 z_1^2, \tag{50}$$

and a constant fluid flow of velocity u along x_1 coordinate axis. In Eq. (50), c_0, a, and b are positive constants. Density of fluid is assumed constant. In such a waveguide, the wave equation (2) for normal modes becomes

$$\partial^2 p/\partial z_1^2 + [\omega^2 c_0^{-2}\beta^2(b - a^2 z_1^2) - q^2]p = 0. \tag{51}$$

By a linear transformation of the independent variable, Eq. (51) reduces to the time-independent Schrödinger equation[34] for the wave functions of stationary states of a quantum-mechanical harmonic oscillator. From the known solution[34] for the harmonic oscillator, one finds mode shape functions and propagation constants

$$f_n^a(z_1) = N_n \exp(-y_n^2/2) H_{|n|-1}(y_n),$$
$$y_n = (\omega a \beta_n/c_0)^{1/2} z_1, \quad n = \pm 1, \pm 2, \ldots, \tag{52}$$

$$q_n^a = \frac{\omega/c_0}{1 - bu^2/c_0^2}\left[\frac{au}{\omega}\left(|n| - \frac{1}{2}\right) - \frac{bu}{c_0} + \frac{n}{|n|}\sqrt{b - \frac{2ac_0}{\omega}\left(|n| - \frac{1}{2}\right) + \frac{a^2u^2}{\omega^2}\left(|n| - \frac{1}{2}\right)^2}\right], \tag{53}$$

where $H_m(y)$ are Hermite polynomials.[34,35] Numerical factors N_n are calculated from the normalization condition (3)

$$N_n = \frac{(\omega a \rho \beta_n^7/\pi c_0)^{1/4}}{\sqrt{2^{|n|-1}(|n|-1)!}}\left[1 + \frac{au}{q_n c_0}\sqrt{\beta_n^2}\left(|n| - \frac{1}{2}\right)\right]^{-1/2}, \tag{54}$$

using properties of Hermite polynomials.[35,36] The quantity $\beta_n = 1 - q_n^a u/\omega$ may be complex. In Eqs. (52)–(54) it is assumed $-\pi < \arg\beta_n < \pi$.

When viewed in the original coordinate system (x, z), the waveguide described above is range dependent. The waveguide axis, where sound speed is at a minimum, is given by the equation $z = z_0(x)$ with $z_0(x) = x\tan\alpha$; sound speed depends on both x and z, and flow velocity has horizontal $u_1 = u\cos\alpha$ and vertical $u_3 = -u\sin\alpha$ components. Comparing Eq. (2) for local modes to Eq. (51), one finds that shape functions and propagation constants of the local modes are given by Eqs. (52)–(54) with $z - z_0(x)$, $a\cos\alpha$, and $u\cos\alpha$ substituted, respectively, for z_1, a, and u. Unlike the shape functions, propagation constants of the local modes do not depend on range.

Consider now mode-coupling equations for the waveguide at hand. According to Eqs. (33) and (34), $E_{nm}^{(2)} \equiv E_{nm}^{(3)} \equiv 0$ due to $\nabla p_0 \equiv 0$. Note that shape functions f_n of the local modes are even functions of $z - z_0(x)$ for odd $|n|$ and odd functions of $z - z_0(x)$ for even $|n|$, while $\partial c^{-2}/\partial x$ is an odd function of $z - z_0(x)$. Then, it follows from Eqs. (30)–(32) that contributions to the coupling coefficients g_{nm} due to range dependence of the sound speed and due to the vertical component of the flow velocity are nonzero when $|n-m| = 1,3,5,\ldots$; $g_{nm} = 0$ when $|n-m| = 0,2,4,\ldots$. Moreover, because medium parameters and the shape functions depend on range only through the combination $z - z_0(x)$, and calculation of g_{nm} involves integration over z, the coupling coefficients prove to be independent of range.

When a given linear combination of normal modes of frequency ω propagates in the rotated waveguide, the exact solution for acoustic pressure is given by

$$p = \sum_n \eta_n f_n^a(z\cos\alpha - x\sin\alpha) \times \exp(iq_n^a x\cos\alpha + iq_n^a z\sin\alpha - i\omega t), \tag{55}$$

where η_n are arbitrarily prescribed complex amplitudes of the normal modes while f_n^a and q_n^a, are calculated from explicit Eqs. (52)–(54). It is straightforward to write down counterparts of Eq. (55) for oscillatory displacement and other components of the state vector Φ. Equation (55) defines the benchmark solution. As usual, it is the real part of p Eq. (55) that has a direct physical meaning.

To solve mode-coupling equations numerically, we first truncate the infinite set of Eqs. (29) and then calculate the coupling coefficients g_{nm} Eq. (30) by numerical integration. Finally, the truncated set of the mode-coupling equations was solved with boundary conditions at $x = 0$ obtained from Eq. (55). The calculations were implemented in the software package MATHEMATICA.[36] Figures 1–3 illustrate results of a comparison of the numerical and benchmark solutions.

Two versions of the truncated Eq. (29) were considered: (i) with ten equations for unknown mode amplitudes F_n, $|n| < 6$ and (ii) with 20 equations for F_n with $|n| < 11$. To emphasize the exact nature of the coupled-mode formulation, parameters of the waveguide are chosen in such a way that mode coupling is significant and flow velocity is not small compared to the sound speed. Specifically, let $\omega/ac_0 = 1$, $u/c_0 = 0.1$, $b = 10$, $\alpha = 0.5$ radian (Figs. 1 and 2). In this case, ratio of the flow velocity to the sound speed on the waveguide axis is $b^{1/2}u/c_0 \approx 0.32$. The rotated waveguide supports ten propagating modes with $\mathrm{Im}\, q_n = 0$; modes with $|n| > 5$ are evanescent, i.e., $\mathrm{Im}\, q_n \neq 0$. When the benchmark solution corresponds to a single propagating mode of the rotated waveguide, exact as well as numerical solutions of the coupled-mode equations predict that, for fixed $z - z_0(x)$, $|p|$ is independent of x. Variation of the acoustic pressure amplitude in a waveguide cross section $x = \mathrm{const}$ is shown in Figs.

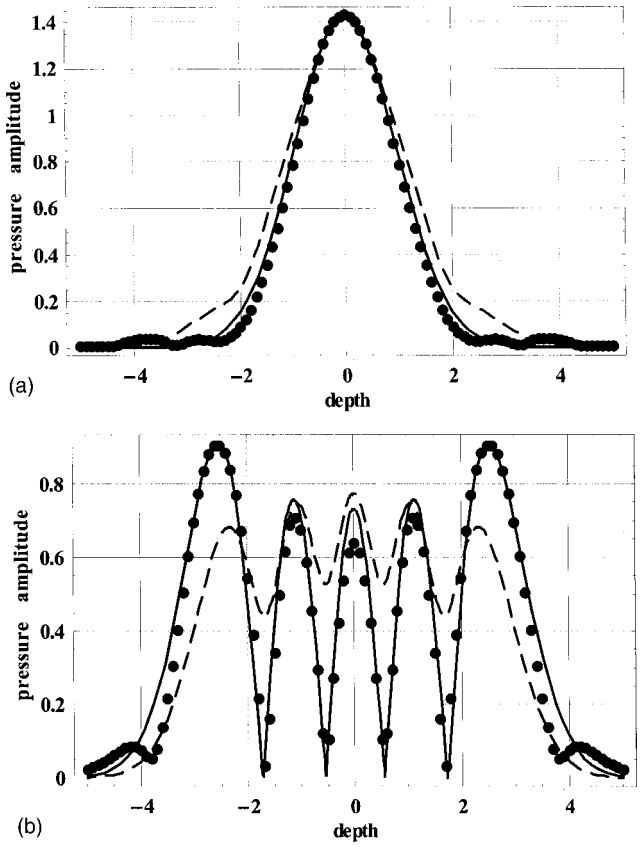


FIG. 1. Acoustic pressure amplitude in the rotated waveguide when a single normal mode of unit amplitude propagates upstream: exact solution (solid lines) and numerical solutions of the set of coupled-mode equations (29) truncated to 10 equations (dashed lines) and 20 equations (dots). The pressure amplitude is shown as a function of the dimensionless depth $\omega c_0^{-1}(z - z_0(x))$ measured from the waveguide's axis $z = z_0(x)$. The normal mode order is $n = -1$ (a) and $n = -5$ (b). Parameters of the waveguide are defined in the text.

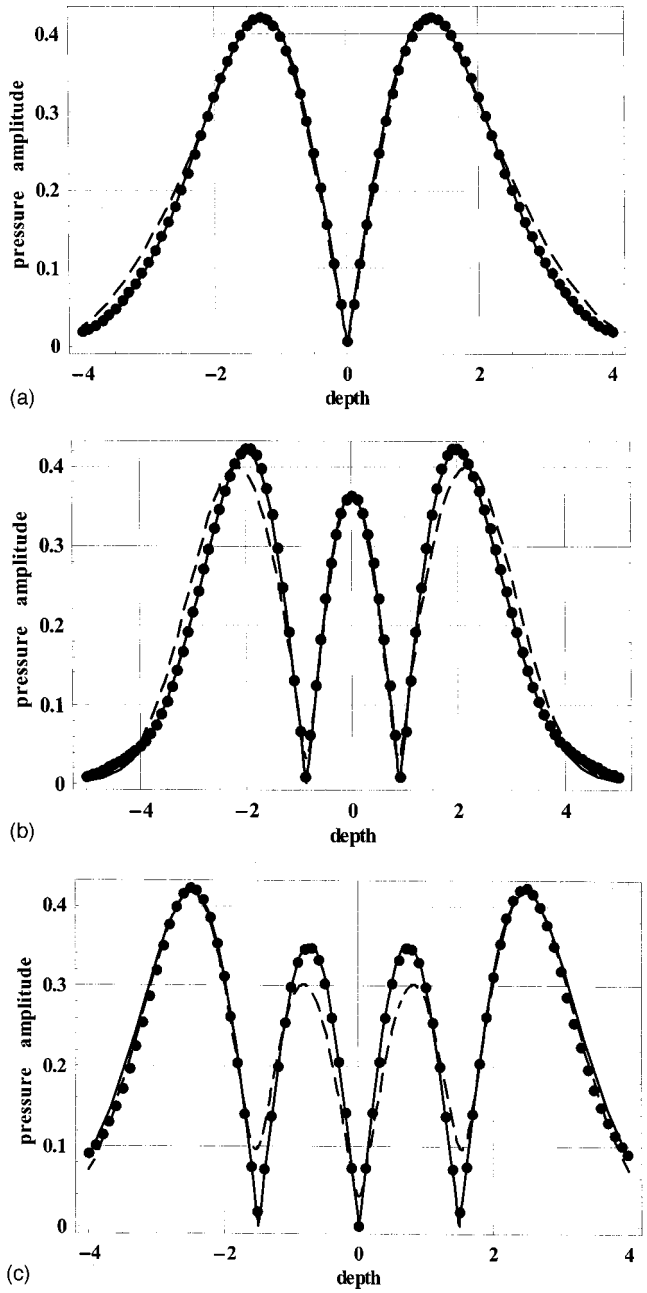


FIG. 2. The same as Fig. 1 for normal modes of the rotated waveguide propagating downstream: $n = 2$ (a); $n = 3$ (b); and $n = 4$ (c).

1 and 2 for different orders of normal modes of the rotated waveguide. Mode coupling is strong, as can be judged from large values of the mode-coupling coefficients or from spectral content of the mode of the rotated waveguide in terms of the local modes. For instance, there are five local modes with excitation coefficients of 20% of the largest excitation coefficient or higher for the mode $n = 5$ and 11 such local modes for the mode $n = -2$ of the rotated waveguide. Despite the strong coupling, even the set of ten coupling equations provides a reasonably accurate model for calculating the amplitude cross sections, especially for lowest-order modes; see Figs. 1 and 2. Overall agreement for modes propagating downstream is better than for modes propagating upstream. As expected, the agreement between the benchmark and numerical solutions improves with increasing size of the set of the mode-coupling equations. With the bigger of the two truncated sets of the coupling equations, there is a good agreement between the numerical results and the benchmark solution for all propagating modes.

As a phase-sensitive measure of correspondence between the numerical and benchmark solutions, consider an instantaneous value of the acoustic pressure at $t = 0$. Let us assume that two modes of the same frequency propagate downstream in the rotated waveguide. Figure 3 corresponds to the case when the first and fifth modes of the rotated waveguide are excited with unit amplitudes. The figure shows a coupled-mode solution and a solution within the adiabatic approximation. Both solution were required to match the benchmark at $x = 0$. Parameters of the waveguide are the same as defined above with the exception of angle α, which is now chosen to be 0.2 radian. Smaller value of α weakens mode coupling and thus benefits the adiabatic approximation. Nevertheless, as illustrated by Fig. 3, the adiabatic approximation fails to accurately reproduce the true range dependence of amplitudes of the local modes and rapidly deviates from the benchmark solution. When mode coupling is taken into account, agreement between the exact and numerical solutions proves to be rather good. Phase errors accumulate with range, though, eventually leading to a sig-

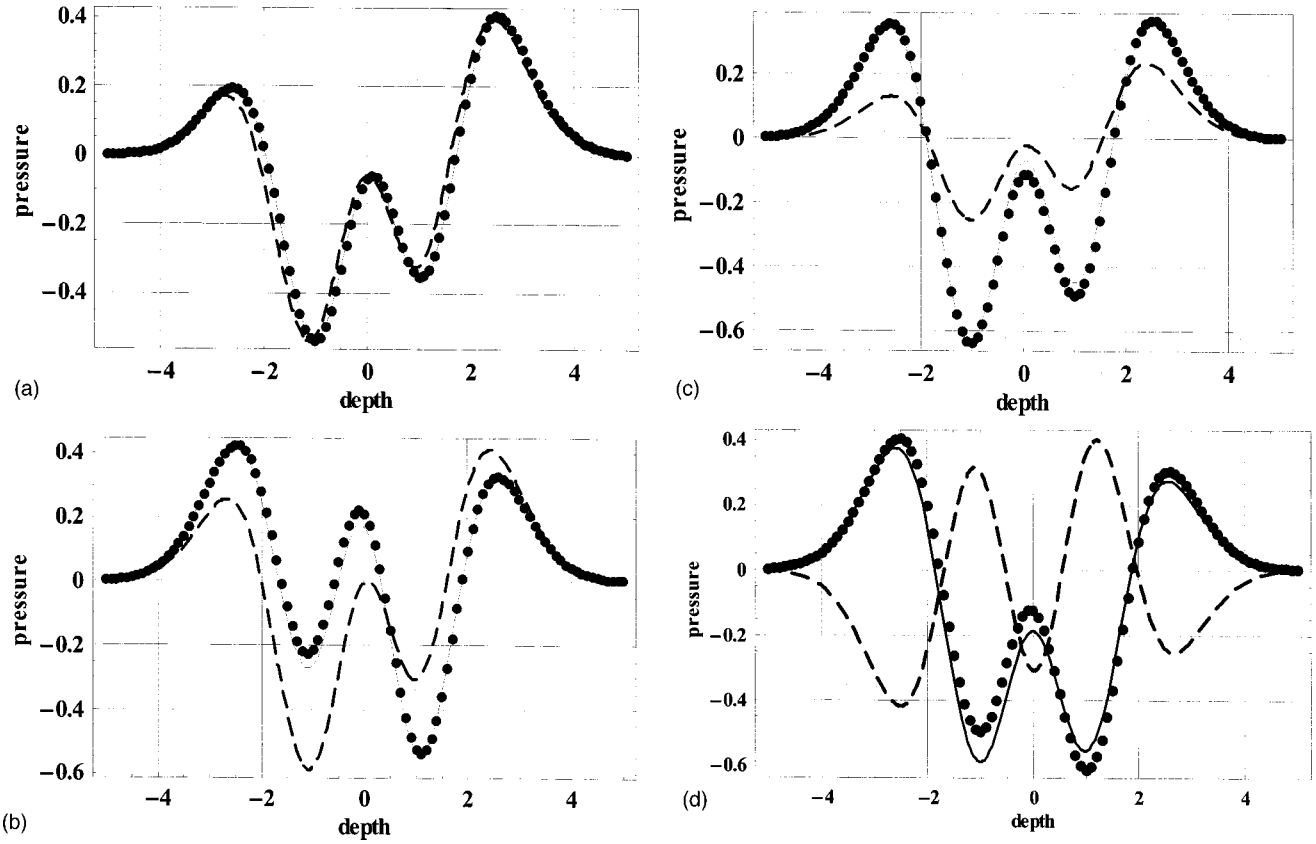


FIG. 3. Acoustic pressure in the rotated waveguide when two normal modes propagate downstream: exact solution (solid line), calculations within the adiabatic approximation (dashed line), and numerical solutions of the set of coupled-mode equations (29) truncated to 20 equations (dots). The pressure at the moment $t=0$ is shown as a function of the dimensionless depth $\omega c_0^{-1}(z-z_0(x))$ measured from the waveguide's axis $z=z_0(x)$, for various values of the dimensionless range $\omega x/c_0$: $\omega x/c_0=1$ (a); $\omega x/c_0=10$ (b); $\omega x/c_0=20$ (c); $\omega x/c_0=50$ (d). Parameters of the waveguide are defined in the text.

nificant discrepancy between the numerical solution and the benchmark. According to Fig. 3(c), the accuracy obtained when the infinite set of coupling equations is truncated to 20 equations is still sufficient to calculate the field at the dimensionless range $\omega x/c_0=50$, which approximately corresponds to 13.57 periods (wavelengths) of interference of the two modes in the benchmark solution.

VII. DISCUSSION

The coupled-mode equations (29) have been derived in Secs. II–IV using the state vector $\mathbf{\Phi}$ (9). After the fact, it is obvious that the coupled-mode equations could have been obtained, without invoking any state vector, by manipulating equations of linear acoustics in terms of acoustic pressure and oscillatory displacement. The rationale for introducing a state vector is, then, to cast a necessarily tedious transformation of the equations of linear acoustics of moving media in the most transparent and intuitive form possible. Of course, the definition Eq. (9) of the state vector is neither the only possible nor the only suitable one. For instance, one can define a state vector as $\mathbf{\Upsilon}=(\varphi,w_1,dw_1/dt,w_3,dw_3/dt)$, where $\varphi=p+\rho u_1 dw_1/dt$. Then, one again obtains the orthogonality relation Eq. (12) and the vector form Eq. (24) of the equations of motion, matrices $[A_\Upsilon]$ and $[B_\Upsilon]$, of course, being different from those corresponding to the state vector $\mathbf{\Phi}$. When using the state vector $\mathbf{\Upsilon}$, details of the calculations differ, but the resulting coupled-mode equations and coupling coefficients are exactly the same as presented in Sec. IV. Actual transformations in terms of $\mathbf{\Upsilon}$ happen to be simpler than in terms of $\mathbf{\Phi}$, mostly because the 5×5 matrix $[B_\Upsilon]$ in the mode orthogonality relation has just four nonzero elements. We, however, prefer a description based on $\mathbf{\Phi}$ because this state vector contains the acoustic pressure among its components and thereby facilitates comparison of coupled-mode equations in moving and motionless media.

In this paper, we assume a continuous range dependence of the environmental parameters that results in continuous mode coupling. This is a realistic model for most problems of interest in environmental acoustics. In numerical simulations, it is often assumed for computational convenience that the waveguide consists of range-independent segments that join at vertical interfaces.[37] Mode coupling occurs then only at these interfaces. For each interface, the amplitudes of outgoing modes can be determined in terms of the amplitudes of incoming modes, using boundary conditions of continuity of appropriate acoustic quantities at the interface. The mode orthogonality relation Eq. (12) obtained in Sec. II is expected to be helpful in calculating mode spectrum transformation at a vertical interface in a moving medium. However, there appears to be a fundamental problem in applying the discrete coupling model to waveguides with flow. Mass conservation requires that the normal component of flow velocity be continuous at the interface (Ref. 3, Sec. 7.3.2). Hence, the depth dependence of the horizontal current velocity component,

u_l, must be the same in any adjacent range-independent segments and, therefore, throughout the waveguide. Strictly speaking, this makes the discrete model inapplicable to waveguides with range-dependent flow velocity. It is not clear whether physically meaningful solutions can be obtained within the discrete-coupling model when the boundary condition of horizontal velocity continuity at vertical interfaces is ignored. No such problems arise in the coupled-mode representation that assumes continuous coupling.

VIII. CONCLUSION

With the local basis chosen as the normal modes of a range-independent waveguide in a moving medium, mode-coupling equations in the presence of currents retain the structure of mode-coupling equations in motionless fluid, with the only difference being the value of the coupling coefficients. In moving media, the range dependence of flow velocity (but not flow velocity itself) contributes to mode coupling, much like the range dependence of sound speed or medium density. To calculate the coupling coefficients in a given waveguide, it is sufficient to know the depth dependence of acoustic pressure in corresponding local modes. The similarity between acoustic mode-coupling equations in moving and in motionless fluids greatly simplifies an extension of the computer codes implementing mode-coupling theory for quiescent media to the case of guided sound propagation through an inhomogeneous flow.

Further research is necessary to extend the mode-coupling theory to the continuous spectrum and to determine mode coupling due to sloping boundaries and interfaces in moving media.

ACKNOWLEDGMENTS

Grateful acknowledgment is given for illuminating discussions with Vladimir A. Borovikov (Benemerita Universidad Autonoma de Puebla, Puebla, Mexico), Michael G. Brown (University of Miami, Miami, FL), and Michael I. Taroudakis (Institute of Applied and Computational Mathematics, Foundation for Research and Technology-Hellas, Heraklion, Greece). The research leading to this paper was originally supported in part by the U.S. Office of Naval Research and the Natural Sciences and Engineering Research Council of Canada, and was completed while the author held a Senior NRC-NOAA/ETL Research Associateship.

[1] O. A. Godin, D. Yu. Mikhin, and D. R. Palmer, ''Monitoring ocean currents in the coastal zone,'' Izv. Atmos. Ocean. Phys. **36**, 131–142 (2000).
[2] P. Elisseeff, and H. Schmidt, ''Acoustic propagation through a low Mach number, stratified flow,'' J. Acoust. Soc. Am. **101**, 1936–1944 (1997).
[3] L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Medium. 2: Point Sources and Bounded Beams*, 2nd ed. (Springer, Berlin, Heidelberg, 1999).
[4] A. D. Pierce, ''Extension of the method of normal modes to sound propagation in an almost-stratified medium,'' J. Acoust. Soc. Am. **37**, 19–27 (1965).
[5] D. M. Milder, ''Ray and wave invariants for SOFAR channel propagation,'' J. Acoust. Soc. Am. **46**, 1259–1263 (1969).
[6] V. V. Shevchenko, ''Irregular acoustic waveguides,'' Sov. Phys. Acoust. **7**, 392–397 (1962).
[7] R. D. Graves, A. Nagl, H. Überall, and G. L. Zarur, ''Range-dependent normal modes in underwater sound propagation: Application to the wedge-shaped ocean,'' J. Acoust. Soc. Am. **58**, 1171–1177 (1975).
[8] S. T. McDaniel, ''Mode coupling due to interaction with the seabed,'' J. Acoust. Soc. Am. **72**, 916–923 (1982).
[9] C. A. Boyles, ''Coupled-mode solution for a cylindrically symmetric ocean waveguide with range and depth-dependent refractive index and a time-varying rough sea surface,'' J. Acoust. Soc. Am. **73**, 800–805 (1983).
[10] M. Hall, ''The effects of variations in sound speed on coupling coefficients between acoustic normal modes in shallow water over a sloping bottom,'' J. Acoust. Soc. Am. **79**, 332–337 (1986).
[11] V. Maupin, ''Surface waves across 2-D structures: A method based on coupled local modes,'' Geophys. J. **33**, 173–185 (1988).
[12] C.-S. Chiu and L. L. Ehret, ''Computation of sound propagation in a three-dimensionally varying ocean: A coupled normal mode approach,'' in *Computational Acoustics*, edited by D. Lee, R. L. Sternberg, and M. H. Schultz, (Elsevier, North-Holland, 1990), Vol. 1, pp. 187–202.
[13] J. A. Fawcett, ''A derivation of the differential equations of coupled mode propagation,'' J. Acoust. Soc. Am. **92**, 290–295 (1992).
[14] B. E. McDonald, M. D. Collins, W. A. Kuperman, and K. D. Heaney, ''Comparison of data and model predictions for Heard island acoustic transmissions,'' J. Acoust. Soc. Am. **96**, 2357–2370 (1994).
[15] G. Gillette, ''Coupled modes in a waveguide with range-dependent rigid basement,'' J. Acoust. Soc. Am. **95**, 187–200 (1994).
[16] M. Razavy, ''Localized and propagating modes in acoustical waveguides with variable cross section,'' J. Acoust. Soc. Am. **95**, 2371–2377 (1994).
[17] Y. Desaubies and K. Dysthe, ''Normal-mode propagation in slowly varying ocean waveguides,'' J. Acoust. Soc. Am. **97**, 933–946 (1995).
[18] A. Voronovich, ''Non-parabolic marching algorithm for sound field calculation in the ocean waveguide,'' J. Comput. Phys. **4**, p. 399–423 (1996).
[19] R. I. Odom, M. Park, J. A. Mercer, R. S. Crosson, and P. Paik, ''Effects of transverse isotropy on modes and mode coupling in shallow water,'' J. Acoust. Soc. Am. **100**, 2079–2092 (1996).
[20] C.-S. Chiu, J. H. Miller, and J. F. Lynch, ''Forward coupled-mode propagation modeling for coastal acoustic tomography,'' J. Acoust. Soc. Am. **99**, 793–802 (1996).
[21] O. A. Godin, ''A note on differential equations of coupled mode propagation in fluids,'' J. Acoust. Soc. Am. **103**, 159–168 (1998).
[22] O. A. Godin, ''On the approximation of a layered moving medium in acoustics,'' Dokl. Akad. Nauk SSSR **316**, 1378–1382 (1991) [Sov. Phys. Dokl. **36**, 168–172 (1991)].
[23] O. A. Godin, ''Reciprocity and energy theorems for waves in a compressible inhomogeneous moving fluid,'' Wave Motion **25**, 143–167 (1997).
[24] In the study of mode-coupling, much like in many other problems in acoustics of moving media (Refs. 23, 25), a sound field in a medium with reversed flow plays the role of a solution to the adjoint boundary-value problem. When $\mathbf{u} \equiv 0$, there is no need to consider an auxiliary, adjoint problem because the problem at hand is then self-adjoint.
[25] O. A. Godin, ''Helmholtz–Kirchhoff integral theorem for waves in the flow of an inhomogeneous compressible fluid,'' Phys. Dokl. **43**, 84–87 (1998).
[26] Alternatively, one can account for gravity in the hydrodynamic equations for an unperturbed medium and neglect effects of gravity in the acoustic Eqs. (14) and (15). Although somewhat inconsistent, such an approach can be justified for waves with frequencies much higher than the Brunt–Väisälä frequency. With such an approach, the analysis of mode coupling is little changed compared to the one presented in the paper.
[27] O. A. Godin, D. Yu. Mikhin, and S. Ya. Molchanov, ''Effective sound speed approximation in the acoustics of moving media,'' Izv. Atmos. Ocean. Phys. **29**, 179–186 (1993).
[28] C. H. E. Warren, ''A note on refraction of sound in moving gas,'' J. Sound Vib. **1**, 175–178 (1964).
[29] R. J. Thompson, ''Ray theory for an inhomogeneous moving medium,'' J. Acoust. Soc. Am. **51**, 1675–1682 (1972).
[30] N. S. Grigor'eva and M. I. Yavor, ''Influence on the sound field in the ocean of a large-scale ocean current that qualitatively alters the nature of guided-wave sound propagation,'' Sov. Phys. Acoust. **32**, 482–485 (1986).
[31] J. S. Robertson, W. L. Siegmann, and M. J. Jacobson, ''Current and current shear effects in the parabolic approximation for underwater sound channels,'' J. Acoust. Soc. Am. **77**, 1768–1780 (1985).
[32] O. A. Godin, ''A wide-angle, energy-conserving parabolic equation for sound in moving medium,'' in *Theoretical and Computational Acoustics '97*, edited by Y.-C. Teng, E.-C. Shang, Y.-H. Pao, M. H. Schultz, and A. D. Pierce (World Scientific, Singapore, 1999), pp. 329–340.
[33] O. A. Godin, ''An effective quiescent medium for sound propagating

through an inhomogeneous moving fluid,'' J. Acoust. Soc. Am. (submitted).
[34] L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3, Quantum Mechanics, Nonrelativistic Theory (Pergamon, New York, 1972), Sec. 23 and Appendix A.
[35] *Handbook of Mathematical Functions with Formulas, Graphs, and Tables*, Appl. Math. Ser., Vol. 55, edited by M. Abramovitz and I. A. Stegun (National Bureau of Standards, Washington, 1964), Chap. 22.
[36] S. Wolfram, *The* MATHEMATICA *Book*, 3rd ed. (Wolfram Research, Champaign, 1996).
[37] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994), Sec. 5.9.

Scattering from impedance gratings and surface wave formation

Wenhao Zhu,[a] Michael R. Stinson, and Gilles A. Daigle
Institute for Microstructural Sciences, National Research Council, Ottawa, Ontario K1A 0R6, Canada

(Received 16 July 2001; revised 13 February 2002; accepted 15 February 2002)

The scattering problem of acoustic plane waves from comb-like impedance gratings on a rigid surface has been investigated in this paper. A rigorous analytic approach for homogeneous plane-wave incidence is presented based on the periodicity of the grating structure, in which the problem was solved as a mixed boundary value problem and the scattered field was represented by the tangent velocity difference across a partition wall of the grating. A singular integral equation has been derived for the tangent velocity difference, which can directly be solved with the Gauss–Chebyshev procedure. The resulting solution consists of a series of Bloch–Floquet waves (plane bulk wave and surface wave modes) with explicit expressions for the expansion coefficients. When the grating period is much less than the incident wavelength ($ka \ll 1$), the grating structure is equivalent to a plane impedance surface and no surface waves can be excited with homogeneous plane-wave incidence. When the grating period is comparable to the incident wavelength, resonance phenomena are predicted under certain conditions and surface waves can form, even with homogeneous plane-wave incidence. The dispersion relation for surface waves has also been examined. The impedance effects of the grating on the reflection and diffraction waves as well as on the dispersion and formation of surface waves have been studied, with the acoustically hard grating being the special case of the general impedance grating. [DOI: 10.1121/1.1468879]

PACS numbers: 43.20.El, 43.20.Fn [LLT]

I. INTRODUCTION

Surface wave propagation above periodic or corrugated surfaces has been observed and studied for a long time[1–4] in the fields of acoustics and electromagnetics. For acoustic air-coupled surface waves, model experiments by Daigle,[5] Daigle and Embleton,[6] and Hutchinson-Howorth and Attenborough[7] have demonstrated the separation of an initial tone burst into a surface wave with a slower speed and the bulk wave with the regular sound speed in air. Subsequent studies by Stinson *et al.*,[8] Daigle *et al.*,[9] and Stinson and Daigle[10] showed that the periodic surface structures could be modeled by an equivalent impedance plane at low frequencies so that analytical methods[11,12] could be applied in calculating the surface and bulk wave fields. An intuitive description of surface-wave formation above periodic surfaces was given by Tolstoy[13] in terms of the energy storing and releasing mechanism between the oscillating fields in the concave regions of the structures and the surface waves right above the structures. With a low-frequency approximation, Tolstoy[14] used a simple effective boundary condition to replace the complicated original one and obtained with point-source incidence the boundary (surface) wave fields above slightly rough surfaces, a more general case of surface geometry. Recently, a modal model method, which was initially developed to model optical or electromagnetic wave scattering by gratings, has been employed by Kelders *et al.*[15] and Lauriks *et al.*[16] to calculate the dispersion relation for acoustic surface waves above a grating. They compared the predicted results by this method with the measured results.

Bulk wave reflection and diffraction by periodic surfaces or gratings have received extensive study from both acoustic and electromagnetic areas for nearly a century, starting from Rayleigh's classical work.[17] Since a complete review on this subject is not intended to be included in this work, we refer the reader to some relevant papers by Lippman,[18] Millar,[19] Waterman,[20] Wirgin,[21] DeSanto,[22,23] and Holford,[24] and references therein.

To the authors' knowledge, existing theoretical treatments do not include the case of gratings with finite acoustic impedance. It seems that except for the geometry difference, the problem of periodic liquid/solid interfaces[25,26] may result in a similar impedance boundary condition as we will discuss here, if one simplifies the interface effect on the field in either medium with an ''equivalent'' impedance boundary. However, such an equivalent impedance boundary condition will be field dependent (nonlocally reacting); moreover, the previous studies on the corrugated interface problem assume small corrugation heights to enable the Rayleigh approximation. Most existing results on surface waves above grating structures have been obtained by making a low-frequency approximation (i.e., the incidence wavelength is much larger than the grating period). Using this approximation, homogeneous plane-wave incidence cannot excite effective surface waves over corrugated surfaces. It is unknown if this conclusion is also applicable to medium- or high-frequency cases. In this paper, the problem of scattering and surface-wave propagation above a comb-like, impedance grating under plane-wave incidence is considered without using the low-frequency approximation. The problem was met when the authors investigated the potential application of acoustic surface waves in the communication area, and it may also have useful applications in other areas.

[a]Electronic mail: wenhao.zhu@nrc.ca

An analytical method is described based on the periodicity of the geometry to reduce the problem to a singular integral equation for the tangent velocity difference crossing a partition wall of the grating. Unlike the modal model method,[15,16] this approach results in a Bloch–Floquet wave expansion with explicit expression of the expansion coefficients. The Gauss–Chebyshev procedure is employed in solving the singular integral equation numerically. It is demonstrated through numerical simulations that homogeneous plane-wave incidences can excite dominant surface-wave propagation over an infinite grating structure under certain conditions. In the following section, the grating scattering formalism is derived for a comb-like grating geometry with an acoustic leakage impedance under plane wave incidence. In Sec. III, a singular integral equation is obtained and solved numerically, from which the scattered bulk waves as well as surface waves can be calculated. Numerical results on scattered bulk waves are given in Sec. IV for the specular reflection and the first-order diffraction. In particular, the low-frequency approximation is examined and a comparison is made to known results. One important issue in equating a grating structure to a flat impedance surface is the ''locally reacting'' property of the equivalent impedance. This is discussed in detail in the section. The surface-wave dispersion is the topic of Sec. V, in which the exact solution is compared with the simplified dispersion equations at low frequencies ($ka \ll 1$). The impedance effects are also described. Finally, the resonant phenomena among surface modes in the grating scattering are predicted through several examples in Sec. VI. It is explained how this can be used in exciting effective surface waves over a grating with plane-wave incidence.

II. PROBLEM FORMULATION AND SOLUTION FOR PLANE-WAVE INCIDENCE

Consider a comb-like structure vertically mounted on a rigid infinite plane as shown in Fig. 1(a). The partition walls, assumed infinitely thin (compared to the wall spacing) and extending infinitely in the y direction, have the same height l and are equally spaced with period a. We assume the partition walls are not perfect reflection plates but, instead, are a kind of porous-plate structure with densely distributed small holes and/or gaps. A continuous acoustic leakage or impedance model is employed to characterize such nonideal partition walls, as shown in Fig. 1(b). In the model, the pressure difference across a wall will result in a nonzero particle velocity through the wall, which is proportional to the pressure difference. Letting ϕ be the velocity potential, P, the pressure, and Z, the leakage-specific impedance normalized by the characteristic impedance (ρc) of air, we have the following boundary condition along a partition wall for the quantities defined on either side of the wall:

$$\left.\frac{\partial \phi}{\partial x}\right|_{+} = \frac{(P_- - P_+)}{\rho c Z} = ik(\phi_+ - \phi_-)/Z, \tag{1a}$$

$$\left.\frac{\partial \phi}{\partial x}\right|_{-} = \left.\frac{\partial \phi}{\partial x}\right|_{+}, \tag{1b}$$

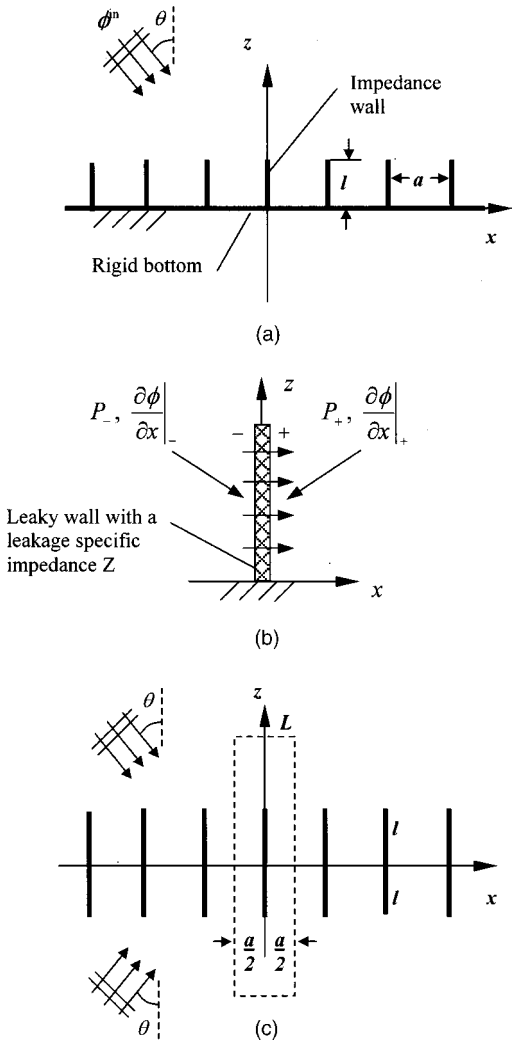


FIG. 1. The geometry of the impedance grating and incident waves: (a) the comb-like grating; (b) the impedance model describing the acoustic leakage; and (c) the equivalent full-space model.

where the ''+,'' ''−'' sides of the wall are defined in Fig. 1(b), k is the wave number in air, and the time factor $\exp(-i\omega t)$ is omitted hereafter. The normal velocity continuity Eq. (1b) comes from the material conservation law. With this model, the material property of the impedance grating is characterized by a (distributed) constant impedance Z, and when it is set to be infinite, Eq. (1) reduces to the boundary condition for acoustically hard walls. The grating structure is insonified by a time-harmonic, homogeneous plane wave with an angle of incidence θ. As the structure is periodic in one dimension (in x), only plane waves incident with wave vectors lying in the $x-z$ plane are considered, and the extension to the case of a general plane-wave incidence is relatively simple. This leads to a two-dimensional scattering problem in the upper-half space.

The problem is equivalent to a full-space boundary value problem with a symmetry plane at $z=0$, where the rigid bottom surface condition is automatically satisfied by using a second incident plane wave, as shown in Fig. 1(c). It will be more convenient to consider the equivalent problem in the full space than the original problem in the upper-half space, when a Fourier transform in z is applied. In either

case, we express the total velocity potential field as the sum of the incident field and the scattered field: [Note: quantities with subscript "*f*" are related to the full space shown in Fig. 1(c)]

$$\phi(x,z)=\phi_f(x,z), \quad z\geqslant 0, \tag{2a}$$

where

$$\phi(x,z)\equiv e^{i\xi_0 x - i\eta_0 z}+\phi^{sc}(x,z), \tag{2b}$$

$$\phi_f(x,z)\equiv \phi_f^{in}(x,z)+\phi_f^{sc}(x,z). \tag{2c}$$

The "incident field" in the full space is given by [Fig. 1(c)]

$$\begin{aligned}\phi_f^{in}(x,z)&=e^{i\xi_0 x - i\eta_0 z}+e^{i\xi_0 x + i\eta_0 z}\\&=2\cos(\eta_0 z)\exp(i\xi_0 x),\end{aligned} \tag{3}$$

where ξ_0 is the incident wave number in the x direction and

$$\eta_0=\sqrt{k^2-\xi_0^2}, \quad k=\omega/c, \tag{4}$$

where ω is the angular frequency and c the sound speed in air. Generally, ξ_0 and η_0 can be complex numbers, but initially, only homogeneous plane waves are considered, so

$$\xi_0=k\sin\theta, \quad \eta_0=k\cos\theta. \tag{5}$$

Then, from Eqs. (1) and (2), the scattered field is determined by

$$\nabla^2\phi_f^{sc}(x,z)+k^2\phi_f^{sc}(x,z)=0, \tag{6}$$

$$\begin{aligned}&\frac{\partial\phi_f^{sc}(x_n^+,z)}{\partial x}-\frac{ik}{Z}[\phi_f^{sc}(x_n^+,z)-\phi_f^{sc}(x_n^-,z)]\\&\quad=-\frac{\partial\phi_f^{in}(x_n,z)}{\partial x}=-2i\xi_0\cos(\eta_0 z)e^{i\xi_0 x_n}, \quad |z|\leqslant l,\end{aligned} \tag{7a}$$

$$\frac{\partial\phi_f^{sc}(x_n^-,z)}{\partial x}=\frac{\partial\phi_f^{sc}(x_n^+,z)}{\partial x}, \quad |z|\leqslant\infty, \tag{7b}$$

where $x_n=na$, $n=0,\pm 1,...,\pm\infty$, are the locations of the walls. Equation (7b) is set to be valid for any z, since the scattered velocity field is still continuous when $|z|>l$. It is seen from Eq. (7) that the boundary geometry is periodic in x and so is the boundary condition, except for the term $\exp(i\xi_0 x)$. This implies that the Bloch–Floquet theory[27] may be applied here and the scattered field should have a term that is in common with the incident field

$$\phi_f^{sc}(x,z)=\psi(x,z)e^{i\xi_0 x}, \quad \psi(x+a,z)=\psi(x,z), \tag{8}$$

so $\psi(x,z)$ has the same periodicity as the grating and satisfies the following governing equations:

$$\nabla^2\psi(x,z)+2i\xi_0\partial\psi(x,z)/\partial x+(k^2-\xi_0^2)\psi(x,z)=0, \tag{9}$$

and

$$\begin{aligned}&\frac{\partial\psi(0^+,z)}{\partial x}+i\xi_0\psi(0^+,z)-\frac{ik}{Z}[\psi(0^+,z)-\psi(0^-,z)]\\&\quad=-2i\xi_0\cos(\eta_0 z), \quad |z|\leqslant l,\end{aligned} \tag{10a}$$

$$\begin{aligned}&\frac{\partial\psi(0^-,z)}{\partial x}+i\xi_0\psi(0^-,z)=\frac{\partial\psi(0^+,z)}{\partial x}+i\xi_0\psi(0^+,z),\\&\qquad |z|\leqslant\infty.\end{aligned} \tag{10b}$$

Now that ψ is periodic, we need only consider here the solution of Eqs. (9) and (10) in the region $|x|\leqslant a/2$ in the $x-z$ plane [see Fig. 1(c)], with the periodic condition

$$\begin{aligned}&\psi\left(-\frac{a}{2},z\right)=\psi\left(\frac{a}{2},z\right),\\&\left.\frac{\partial\psi(x,z)}{\partial x}\right|_{x=-a/2}=\left.\frac{\partial\psi(x,z)}{\partial x}\right|_{x=a/2}.\end{aligned} \tag{11}$$

Applying the Fourier transforms on the variable z in Eq. (9), a general solution for $|x|\leqslant a/2$ can be obtained

$$\begin{aligned}&\overline{\psi}(x,\beta)\\&\quad=\begin{cases}A_1(\beta)\exp(\gamma_1 x)+A_2(\beta)\exp(\gamma_2 x), & -a/2\leqslant x\leqslant 0,\\ B_1(\beta)\exp(\gamma_1 x)+B_2(\beta)\exp(\gamma_2 x), & 0\leqslant x\leqslant a/2,\end{cases}\end{aligned} \tag{12}$$

where

$$\overline{\psi}(x,\beta)=\int_{-\infty}^{\infty}\psi(x,z)e^{-i\beta z}\,dz, \tag{13}$$

$$\gamma_{1,2}=-i\xi_0\pm\sqrt{\beta^2-k^2}=-i\xi_0\pm\gamma. \tag{14}$$

By using the periodic condition (11), we have

$$\overline{\psi}(x,\beta)=\begin{cases}A_1(\beta)\exp(\gamma_1 x)+A_2(\beta)\exp(\gamma_2 x),\\ \quad -a/2\leqslant x\leqslant 0,\\ A_1(\beta)\exp(\gamma_1(x-a))+A_2(\beta)\exp(\gamma_2(x-a)),\\ \quad 0\leqslant x\leqslant a/2.\end{cases} \tag{15}$$

Then, applying the continuity condition (10b) results in

$$\begin{aligned}&A_1(\beta)(\gamma_1+i\xi_0)[e^{-\gamma_1 a}-1]+A_2(\beta)(\gamma_2+i\xi_0)\\&\quad\times[e^{-\gamma_2 a}-1]=0.\end{aligned} \tag{16}$$

Let

$$A_1(\beta)=\tfrac{1}{2}[A(\beta)]/[\exp(-\gamma_1 a)-1],$$

$$A_2(\beta)=\tfrac{1}{2}[A(\beta)]/[\exp(-\gamma_2 a)-1],$$

then from Eq. (14), Eq. (16) is automatically satisfied and Eq. (15) becomes

$$\overline{\psi}(x,\beta)=\begin{cases}\tfrac{1}{2}A(\beta)e^{-i\xi_0 x}F_1(x,\beta), & -a/2\leqslant x\leqslant 0,\\ \tfrac{1}{2}A(\beta)e^{-i\xi_0 x}F_2(x,\beta), & 0\leqslant x\leqslant a/2,\end{cases} \tag{17}$$

where

$$F_1(x,\beta)=\frac{\cosh[(x+a)\gamma]-\cosh(x\gamma)e^{-i\xi_0 a}}{\cos(a\xi_0)-\cosh(a\gamma)}, \tag{18a}$$

$$F_2(x,\beta)=\frac{\cosh(x\gamma)e^{i\xi_0 a}-\cosh[(x-a)\gamma]}{\cos(a\xi_0)-\cosh(a\gamma)}. \tag{18b}$$

Before applying the boundary condition (7a) or (10a) to determine $A(\beta)$, we introduce a new unknown function $f(z)$ as follows:

$$f(z)\equiv \frac{\partial \phi_f^{sc}(0^+,z)}{\partial z}-\frac{\partial \phi_f^{sc}(0^-,z)}{\partial z}=\frac{\partial}{\partial z}(\Delta\phi_f^{sc})|_{x=0}. \tag{19}$$

Then, $f(z)=0$ when $|z|>l$, and $f(z)$ is unknown when $|z|\leqslant l$. The Fourier transform of $f(z)$ gives

$$\int_{-l}^{l} f(z)e^{-i\beta z}\,dz=\int_{-\infty}^{\infty} f(z)e^{-i\beta z}\,dz$$
$$=i\beta[\overline{\psi}(0^+,\beta)-\overline{\psi}(0^-,\beta)]=i\beta A(\beta), \tag{20}$$

where Eq. (17) has been used in deriving the last equation. Thus, the unknown $A(\beta)$ can be related to $f(z)$ through

$$A(\beta)=\frac{1}{i\beta}\int_{-l}^{l} f(z)e^{-i\beta z}\,dz. \tag{21}$$

The function $f(z)$ has a physical meaning. From Eq. (19) and the incident field continuity, it represents the tangent velocity jump across any impedance wall, except for a trivial factor $\exp(i\xi_0 na)$. In terms of $f(z)$ and by applying the inverse Fourier transform, we have

$$\psi(x,z)e^{i\xi_0 x}=\frac{1}{4\pi i}\int_{-l}^{l} f(t)\int_{-\infty}^{\infty} F(x,\omega)\frac{e^{i\beta(z-t)}}{\beta}\,d\beta\,dt, \tag{22a}$$

where

$$F(x,\beta)=\begin{cases} F_1(x,\beta), & na-a/2\leqslant x\leqslant na,\\ F_2(x,\beta), & na\leqslant x\leqslant na+a/2,\end{cases} \tag{22b}$$

and where $F_{1,2}(x,\beta)$ are given by Eq. (18). A solution of Eq. (6) represented by Eq. (22a) satisfies the pseudoperiodic condition Eq. (8) and the boundary condition Eq. (7b), but generally is not symmetric about $z=0$. The symmetric scattered field ϕ_f^{sc}, which is required since the incident field Eq. (3) for the full space [Fig. 1(c)] is symmetric about $z=0$, can be obtained by simply choosing the even part of Eq. (22a) with respect to z

$$\phi_f^{sc}(x,z)=[\psi(x,z)+\psi(x,-z)]e^{i\xi_0 x}$$
$$=\frac{1}{4\pi i}\int_{0}^{l} f(t)\int_{-\infty}^{\infty} F(x,\omega)$$
$$\times\frac{e^{i\beta(z-t)}-e^{i\beta(z+t)}+e^{-i\beta(z+t)}-e^{-i\beta(z-t)}}{\beta}\,d\beta\,dt, \tag{23}$$

where use has been made of the fact that $f(z)$ is odd in $[-l,l]$ as ϕ_f^{sc} is even. The four inverse Fourier transforms (inner integrals) in Eq. (23) can be completed by using a contour integration and the residue theorem in the usual way. The results are given in Appendix A. Then, from Eq. (A11), we obtain

$$\phi_f^{sc}=\phi_{f1}^{sc}(x,z)$$
$$=\int_0^l f(t)\left\{\sum_{m=-\infty}^{\infty}\frac{\alpha_m}{ia\beta_m^2}e^{i\alpha_m x+i\beta_m z}[(-2i)\sin(\beta_m t)]\right\}dt,$$
$$z\geqslant l, \tag{24a}$$

and

$$\phi_f^{sc}=\phi_{f2}^{sc}(x,z)$$
$$=\int_0^l f(t)\left\{\sum_{m=-\infty}^{\infty}\frac{\alpha_m}{ia\beta_m^2}e^{i\alpha_m x}[\mathrm{sgn}(z-t)e^{i\beta_m|z-t|}\right.$$
$$\left.-e^{i\beta_m(z+t)}]+\frac{1}{2}[\mathrm{sgn}(z-t)-1]p(x)\right\}dt,$$
$$0\leqslant z\leqslant l, \tag{24b}$$

where $\mathrm{sgn}(u)$ is the signum function, $p(x)$ is given by Eq. (A7), and α_m and β_m are given [see Eq. (A2)] by

$$\alpha_m=\xi_0+\frac{2\pi m}{a},\quad \beta_m=\begin{cases}\sqrt{k^2-\alpha_m^2}, & k^2>\alpha_m^2\\ i\sqrt{\alpha_m^2-k^2}, & k^2<\alpha_m^2\end{cases},$$
$$m=\pm 1,\ldots,\pm\infty. \tag{25}$$

It is obvious from Eqs. (24a) and (24b) that the two solutions defined in the two regions of $z>l$ and $0<z<l$ are continuous at $z=l$

$$\phi_{f1}^{sc}(x,l)=\phi_{f2}^{sc}(x,l).$$

However, the continuity of their derivatives about the z coordinate is not as straightforward. We first write the derivations of Eqs. (24a) and (24b) with respect to z

$$\frac{\partial \phi_{f1}^{sc}(x,z)}{\partial z}$$
$$=\int_0^l f(t)\left\{\sum_{m=-\infty}^{\infty}\frac{\alpha_m}{ia\beta_m^2}e^{i\alpha_m x+i\beta_m z}[(2\beta_m)\sin(\beta_m t)]\right\}dt,$$
$$z\geqslant l, \tag{26a}$$

and

$$\frac{\partial \phi_{f2}^{sc}(x,z)}{\partial z}=\int_0^l f(t)\left\{\sum_{m=-\infty}^{\infty}\frac{\alpha_m}{ia\beta_m^2}e^{i\alpha_m x}(i\beta_m)[e^{i\beta_m|z-t|}\right.$$
$$\left.-e^{i\beta_m(z+t)}]\right\}dt+f(z)\left[\sum_{m=-\infty}^{\infty}\frac{2\alpha_m}{ia\beta_m^2}e^{i\alpha_m x}\right.$$
$$\left.+p(x)\right],\quad z\leqslant l. \tag{26b}$$

As seen from Eqs. (26a) and (26b), the derivatives of the fields will be continuous if and only if the content in the bracket of the second term in Eq. (26b) is equal to zero, i.e.,

$$\sum_{m=-\infty}^{\infty} \frac{2\alpha_m}{ia\beta_m^2} e^{i\alpha_m x} = -p(x), \tag{27}$$

where $p(x)$ is given by Eq. (A7). This has been proven in Appendix B, with the aid of Poisson's summation formula. Therefore, the derivatives with respect to z are also continuous at $z=l$

$$\begin{aligned}\frac{\partial \phi_{f2}^{sc}(x,l)}{\partial z} &= \sum_{m=-\infty}^{\infty} \frac{\alpha_m}{ia\beta_m^2} e^{i\alpha_m x}(i\beta_m) \\ &\quad \times \int_0^l f(t)[e^{i\beta_m(l-t)} - e^{i\beta_m(l+t)}]dt \\ &= \frac{\partial \phi_{f1}^{sc}(x,l)}{\partial z}.\end{aligned}$$

Now, applying the impedance boundary condition (7a) or (10a) to Eq. (24b) and considering the symmetry of the right-hand-side term of Eq. (7a) or (10a) about $z=0$, we obtain an integral equation for $f(t)$ in the upper-half region of $|z| \leqslant l$

$$\begin{aligned}&\int_0^l f(t)\left\{ \sum_{m=-\infty}^{\infty} \frac{\alpha_m^2}{a\beta_m^2}[\mathrm{sgn}(z-t)e^{i\beta_m|z-t|} - e^{i\beta_m(z+t)}] \right. \\ &\left. -[\mathrm{sgn}(z-t)-1]\left[\frac{(k/2)\sin(ak)}{\cos(a\xi_0)-\cos(ak)} + \frac{ik}{Z}\right]\right\} dt \\ &= -\frac{\partial \phi_f^{in}(0,z)}{\partial x} = -2i\xi_0 \cos(\eta_0 z), \quad 0 \leqslant z \leqslant l. \end{aligned} \tag{28}$$

Once $f(t)$ is obtained, the fields above the grating and inside the slots can be determined by Eqs. (24a) and (24b), respectively. Here, we are interested in the scattered field above the grating in the original upper-half space [Fig. 1(a)]. From Eqs. (2) and (24a), this can be expressed in the form of the Bloch–Floquet expansion[24] by exchanging the order of the integral and the series in Eq. (24a)

$$\begin{aligned}\phi^{sc}(x,z) &= \phi(x,z) - e^{i\xi_0 x - i\eta_0 z} \\ &= (1+\Phi_0)e^{i\xi_0 x + i\eta_0 z} \\ &\quad + \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \Phi_m e^{i\alpha_m x + i\beta_m z}, \quad z \geqslant l, \end{aligned} \tag{29}$$

where

$$\Phi_m = -\frac{2\alpha_m}{a\beta_m^2}\int_0^l f(t)\sin(\beta_m t)dt. \tag{30}$$

The scattered field inside the slots ($0<z<l$) can also be expressed by a similar expansion from Eq. (24b). In Eqs. (29) and (30), for a given homogeneous plane-wave incidence and the grating period a, we generally have

$$\beta_m = \begin{cases} \sqrt{k^2-\alpha_m^2}, & -M_1 \leqslant m \leqslant M_2, \\ i\sqrt{\alpha_m^2-k^2}, & -M_1 > m, \quad \text{or,} \quad m > M_2, \end{cases} \tag{31}$$

with

$$M_1 = \left[\frac{a}{\lambda}(1+\sin\theta)\right], \quad M_2 = \left[\frac{a}{\lambda}(1-\sin\theta)\right], \tag{32}$$

where λ is the wavelength of the incident wave and $[s]$ denotes the largest integer in s. Therefore, the scattered field consists of (i) plane bulk waves (positive real β_m), which propagate away from the grating; we call them "diffraction modes," and (ii) evanescent waves decaying in the z direction (positive imaginary β_m), which propagate in a layer immediately above the grating and are usually referred to as "space harmonics" or "surface modes." For a given grating structure, a mode of order m can be either a diffraction mode or a surface mode, depending on the incident field and frequency. The scattered far field is a linear combination of all possible (finite) diffraction modes. As will be seen, the surface modes typically have much smaller amplitude than the diffraction modes, except when certain conditions are met. It is also noted that among the diffraction modes, the first term in Eq. (29) (the zero-order mode) always exists with homogeneous plane-wave incidence and is called the "reflection wave" hereafter since it is along the specular-reflection direction.

The Bloch–Floquet expansion (29) has been called the "Rayleigh expansion" in the literature.[18–24] In the original Rayleigh procedure, a field expansion equivalent to Eq. (29) was assumed valid in both regions $0<z<l$ and $z>l$ (Rayleigh hypothesis), with unknown expansion coefficients. Here, two regional solutions are obtained and the mode amplitudes Φ_m in both regions are given explicitly, though with a function to be determined by an integral equation. It is also noted that in deriving Eqs. (24), (28), and (29), we did not actually specify the incidence field being a homogeneous plane wave. In other word, these equations are still valid for inhomogeneous plane wave incidences, as long as the incident fields have the pseudoperiodicity as defined by Eq. (8).

III. THE SINGULAR INTEGRAL EQUATION

After checking the series in the integral equation (28) it was found that the series converges slowly at $z=t$, since $\alpha_m/\beta_m \to -i$ as $m \to \pm\infty$. By separating the slowly convergent parts from the series, we will obtain a singular term from the kernel and turn Eq. (28) into a Cauchy-type singular integral equation. The energy conservation law is also formulated in this section in order to verify both the solution formalism adopted in the preceding section and the numerical algorithm to be described next.

A. Cauchy-type singular integral equation

First, the series in the kernel of Eq. (28) can be rewritten as follows:

$$\sum_{m=-\infty}^{\infty}\frac{\alpha_m^2}{\beta_m^2}[\mathrm{sgn}(z-t)e^{igb_m|z-t|}-e^{i\beta_m(z+t)}]$$

$$=\sum_{m=1}^{\infty}\left\{\frac{\alpha_m^2}{\beta_m^2}[\mathrm{sgn}(z-t)e^{i\beta_m|z-t|}-e^{i\beta_m(z+t)}]+\frac{\alpha_{-m}^2}{\beta_{-m}^2}[\mathrm{sgn}(z-t)e^{i\beta_{-m}|z-t|}-e^{i\beta_{-m}(z+t)}]+2[\mathrm{sgn}(z-t)e^{-m\nu|z-t|}\right.$$

$$\left.-e^{-m\nu(z+t)}]\right\}+\frac{\xi_0^2}{\eta_0^2}[\mathrm{sgn}(z-t)e^{i\eta_0|z-t|}-e^{i\eta_0(z+t)}]-\left[\frac{2\,\mathrm{sgn}(z-t)}{e^{\nu|z-t|}-1}-\frac{2}{e^{\nu(z+t)}-1}\right]+\left[\frac{2}{\nu(z-t)}-\frac{2}{\nu(z+t)}\right]-\frac{4t}{\nu(t^2-z^2)},\tag{33}$$

where use has been made of

$$\sum_{m=1}^{\infty}e^{-m\nu|z-t|}=\frac{1}{e^{\nu|z-t|}-1},\quad \nu=2\pi/a.\tag{34}$$

Then, substituting Eq. (33) into Eq. (28), we obtain

$$\frac{1}{\pi}\int_0^l\frac{2tf(t)}{t^2-z^2}dt+\frac{1}{a}\int_0^l f(t)[K_0(z-t)+K_1(z-t)]dt$$

$$=-i2\xi_0\cos(\eta_0 z),\quad 0\leqslant z\leqslant l,\tag{35}$$

where

$$K_0(z-t)=\frac{\xi_0^2}{\eta_0^2}[\mathrm{sgn}(z-t)e^{i\eta_0|z-t|}-e^{i\eta_0(z+t)}]$$

$$-[\mathrm{sgn}(z-t)-1]\left[\frac{(ak/2)\sin(ak)}{\cos(a\xi_0)-\cos(ak)}+\frac{ik}{Z}\right]$$

$$-\left[\frac{2\,\mathrm{sgn}(z-t)}{e^{\nu|z-t|}-1}-\frac{2}{e^{\nu(z+t)}-1}\right]$$

$$+\left[\frac{2}{\nu(z-t)}-\frac{2}{\nu(z+t)}\right],\tag{36a}$$

$$K_1(z-t)=\sum_{m=1}^{\infty}\left\{\frac{\alpha_m^2}{\beta_m^2}[\mathrm{sgn}(z-t)e^{i\beta_m|z-t|}-e^{i\beta_m(z+t)}]\right.$$

$$+\frac{\alpha_{-m}^2}{\beta_{-m}^2}[\mathrm{sgn}(z-t)e^{i\beta_{-m}|z-t|}-e^{i\beta_{-m}(z+t)}]$$

$$\left.+2[\mathrm{sgn}(z-t)e^{-m\nu|z-t|}-e^{-m\nu(z+t)}]\right\},\tag{36b}$$

and use has been made of the identity $\mathrm{sgn}(z-t)/|z-t|=1/(z-t)$.

Now, the kernel functions $K_{0,1}(z-t)$ are no longer singular as $t\to z$. Equation (35) is a Cauchy-type singular integral equation of the first kind that has been widely studied.[28] A convenient numerical approach to solving Eq. (35) is the Gauss–Chebyshev procedure.[29] Here, we only give the final linear equations after using the procedure without the detailed derivations.

As is well known in edge-scattering problems, at an edge tip, the scattered velocity field has an $r^{-1/2}$ singularity.[30] We set

$$f(t)=(1-t^2/l^2)^{-1/2}\sum_{k=1}^{\infty}C_kT_{2k-1}(t)\tag{37}$$

in Eq. (35) to obtain

$$\sum_{k=1}^{\infty}\left(a_{jk}+\frac{\pi}{4}\delta_{jk}\right)C_k=b_j,\quad j=1,2,...,\infty,\tag{38}$$

where

$$a_{jk}=\frac{1}{al}\int_0^l\int_0^l[K_0(z-t)+K_1(z-t)]$$

$$\times\frac{\sqrt{1-z^2/l^2}}{\sqrt{1-t^2/l^2}}T_{2k-1}(t)U_{2j-2}(z)dt\,dz,\tag{39a}$$

$$b_j=-2i\xi_0\int_0^l\cos(\eta_0 z)\sqrt{1-z^2/l^2}U_{2j-2}(z)dz,\tag{39b}$$

and

$$T_k(t)=\cos[k\arccos(t/l)],\tag{40a}$$

$$U_j(z)=\sin[(j+1)\arccos(z/l)]/\sin[\arccos(z/l)],\tag{40b}$$

are the first and second Chebyshev polynomials, respectively, and δ_{jk} is the Kronecker delta. The system of linear equations (38) is numerically stable because of the dominant diagonal elements (the δ_{jk} term), resulting from the singularity separation process. It will be truncated and solved for the coefficients C_k. The scattered wave amplitudes are then calculated from Eqs. (29) and (30). In the actual calculation, the series in Eq. (36b) is found to be nearly convergent after 10–15 terms (it is truncated at 15 terms). The series in expression (37) shows almost no difference in the numerical test results after 25 terms (it is truncated at 26 terms).

B. Power transmission and energy conservation

One of the commonly used approaches for assessing the accuracy of the numerical solution for wave propagation problems is the energy conservation law. Consider the rectangular volume consisting of the region shown in Fig. 1(c) (dashed lines) with a unit extension in the y direction. Because of the symmetry of the problem about $z=0$, there is no energy exchange between the upper-half and the lower-half portions of the region. The rate of energy transmission into the volume for time-harmonic fields can be expressed as

$$E=-\frac{1}{2}\rho\omega\,\mathrm{Im}\left[\int_S\left(\frac{\partial\phi}{\partial n}\right)\phi^*\,ds\right], \tag{41}$$

where a "$*$" denotes complex conjugation. S consists of the boundary surfaces $x=a/2$ with $0\leqslant z\leqslant L$, $z=L$ with $-a/2\leqslant x\leqslant a/2$, and $x=-a/2$ with $0\leqslant z\leqslant L$ (the boundary $z=0$ is not included because of the symmetry). Since ϕ is periodic in x with the period a except for the factor $\exp(i\xi_0 x)$, we have

$$\left[\left(\frac{\partial\phi}{\partial x}\right)\phi^*\right]\Bigg|_{x=a/2}=\left[\left(\frac{\partial\phi}{\partial x}\right)\phi^*\right]\Bigg|_{x=-a/2}.$$

Thus, the integrals on $x=a/2$ and $x=-a/2$ will cancel each other. Finally, Eq. (41) is reduced to an integral on $z=L$

$$\begin{aligned}E&=-\frac{\rho\omega}{2}\mathrm{Im}\left[\int_{-a/2}^{a/2}(\phi^{in}+\phi^{sc})^*\frac{\partial}{\partial z}(\phi^{in}+\phi^{sc})\,dx\right]\\&=-\frac{\rho\omega}{2}a\left[\eta_0|1+\Phi_0|^2+\sum_{\substack{m=-M_1\\m\neq 0}}^{M_2}\beta_m|\Phi_m|^2-\eta_0\right],\end{aligned} \tag{42}$$

in which only the diffraction mode contributions are taken into account as the surface modes vanish when L is large enough. When the leakage impedance Z of the partition walls is purely reactive or if the walls are rigid, i.e., no energy loss, the total energy in the volume should be conserved and the net energy flux into the volume should be zero. Therefore, the following relation must hold:

$$|1+\Phi_0|^2+\sum_{\substack{m=-M_1\\m\neq 0}}^{M_2}\frac{\beta_m}{\eta_0}|\Phi_m|^2=\sum_{m=-M_1}^{M_2}E_m=1, \tag{43}$$

where

$$E_0=|1+\Phi_0|^2,\quad E_m=\frac{\beta_m}{\eta_0}|\Phi_m|^2, \tag{44}$$

are defined as the reflection energy efficiency and diffraction energy efficiency for each order m (note that the incident wave has unit amplitude). The condition (43), though not sufficient, is necessary for a correct solution and will be checked frequently in the following numerical results.

IV. NUMERICAL RESULTS ON SCATTERED BULK WAVES AND EQUIVALENT IMPEDANCE SURFACE

Since there are many theoretical and experimental results on bulk wave reflection and diffraction from acoustically hard or -soft periodic surfaces, only two cases are considered here, i.e., the specular reflection alone and then the reflection plus the first-order diffraction. Attention is focused on the equivalent impedance surface and the related "local reaction" condition, the leakage impedance effects of the grating on the reflection and diffraction waves, and other issues related to the surface waves.

A. The reflection wave: Equivalent impedance surface

For homogeneous plane-wave incidence, if the frequency is low enough that

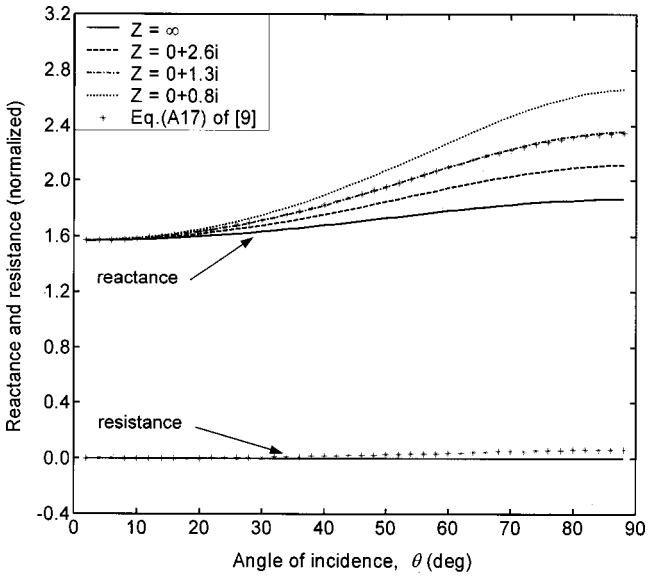


FIG. 2. Equivalent impedance versus angle of incidence ($f=1200$ Hz, $a=0.015$ m, $l=0.0256$ m).

$$\lambda/a>1+\sin\theta, \tag{45}$$

then from Eq. (32), there is only one diffraction mode that can exist in the scattered field expansion (29), i.e., the reflection wave ($m=0$). The far field can be determined by neglecting all the surface modes (space harmonics) in Eq. (29)

$$\begin{aligned}\phi_{\mathrm{far}}(x,z)&=\exp(i\xi_0 x-i\eta_0 z)\\&\quad+(1+\Phi_0)\exp(i\xi_0 x+i\eta_0 z).\end{aligned} \tag{46}$$

The first term is the incident plane wave; the second term, the reflection wave, consists of the mirror reflection from the rigid bottom and a term (Φ_0) due to the existence of the impedance grating structure.

In view of the similarity of this case to the impedance surface reflection, we expect that an equivalent impedance value can be defined for the grating. We assume a virtual impedance plane is located at $z=l$, with a constant impedance Z_{eq} (normalized by the impedance of air ρc). Then, the impedance boundary condition will be

$$\left[\frac{\partial\phi}{\partial z}+\frac{ik}{Z_{\mathrm{eq}}}\phi\right]\Bigg|_{z=l}=0. \tag{47}$$

Substituting Eq. (46) into the above equation yields

$$Z_{\mathrm{eq}}=\left(\frac{k}{\eta_0}\right)\frac{1+R_0e^{i2\eta_0 l}}{1-R_0e^{i2\eta_0 l}},\quad R_0=1+\Phi_0. \tag{48}$$

The equivalent impedance has been calculated as a function of the angle of incidence and is shown in Fig. 2 for four different impedance values of the partition walls ($Z=\infty$, $2.6i$, $1.3i$, and $0.8i$). The calculation is made using $c=340$ m/s, $f=1200$ Hz, $l=0.0256$ m, and $a=0.015$ m. The figure shows that the reactance of Z_{eq} increases as Z decreases (more acoustic leakage) at large angles of incidence. The resistive part of Z_{eq} is zero because Z is assumed to be purely reactive.

Daigle *et al.*[9] suggested a formula for calculating the equivalent impedance of a lattice structure with a lumped,

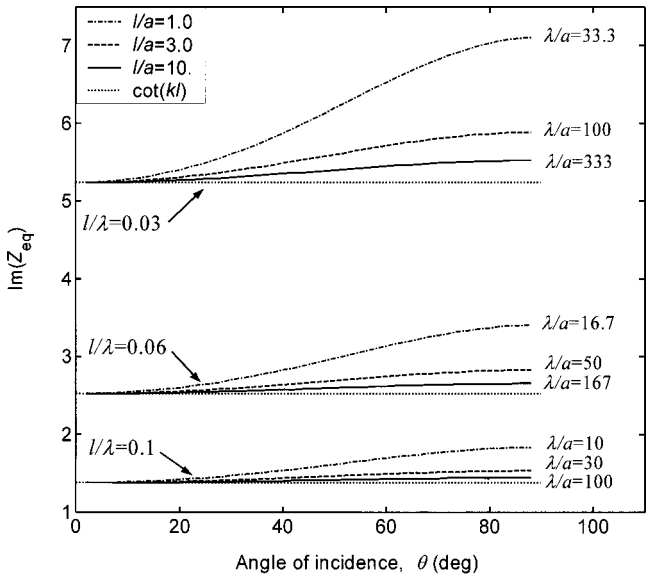


FIG. 3. Imaginary component of the equivalent impedance versus θ for three l/a ratios and three l/λ ratios, respectively ($Z=\infty$).

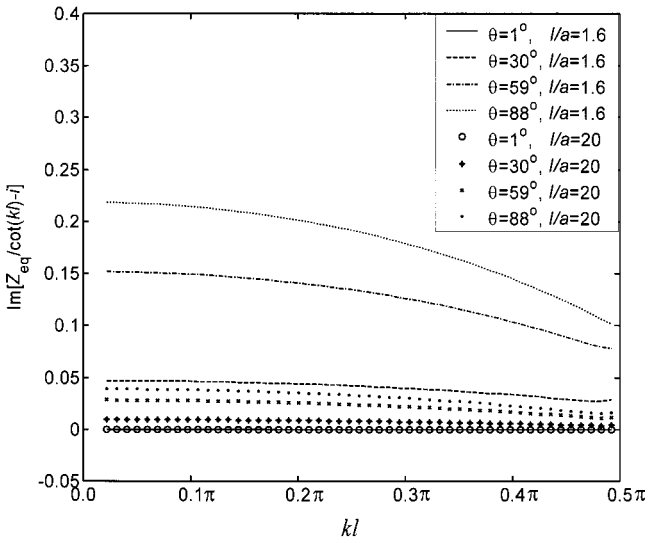


FIG. 4. Relative difference between Eqs. (48) and (49) varying with kl for two l/a ratios: 1.6 and 20, and several angles of incidence ($Z=\infty$).

leakage impedance on each partition wall. By setting the bottom surface as rigid and using the same grating parameters as here, their Eq. (A17) gives an equivalent impedance that is plotted in Fig. 2 as well for comparison. It should be pointed out that the leakage impedance values used for the walls of the gratings are different for the two cases. Here, Z is a continually distributed parameter, whereas in Ref. 9 it is a lumped parameter (including a nonzero resistance) defined at the center of each wall. Nevertheless, the two procedures predict a similar increasing trend of the equivalent impedance (reactance) with the angle of incidence, their reactance curve essentially matching our calculation for $Z=1.3i$.

In general, the equivalent impedance of a grating structure depends on the angle of incidence (as seen in Fig. 2). Therefore, the impedance at a given frequency depends on the acoustic field above the structure and the surface is nonlocally reacting. However, real material surfaces are often treated as locally reacting, with their impedance values independent of the acoustic field. It is of interest to find the conditions under which the grating structure can be equivalent to a flat, locally reacting impedance surface. For simplicity, only rigid gratings are considered in the following. We first examine a special case of Eq. (48), that of normal incidence. From Eqs. (5) and (28), we have $\xi_0=0$, $\eta_0=k$, and $f(z)=0$, therefore $\Phi_0=0$ and Eq. (48) reduces to the well-known results[2,4] for normal incidence

$$Z_{\mathrm{eq}}=i\cot(kl). \tag{49}$$

Equation (49) is generally assumed to be applicable as well to oblique incidence[2,13] when $ka\ll 1$. The validation of this assumption has been tested by calculating Z_{eq} from both Eq. (48) and Eq. (49). The dependence on angle of incidence is shown in Fig. 3, for various ratios l/a and l/λ. When $l/a\gg 1$ and $\lambda/a\gg 1$, the grating structure exhibits nearly locally reacting behavior with the largest deviation occurring at grazing incidence, and Eq. (49) is a good approximation for the equivalent impedance. We also calculate the relative deviation, the difference between Eqs. (48) and (49) divided by Eq. (49), versus the normalized frequency kl for two l/a ratios (1.6 and 20) and several angles of incidence. The results are given in Fig. 4. While significant drops in the relative deviation are observed in the figure as the l/a ratio changes from 1.6 to 20, the relative deviation remains nearly constant for low kl values and for the larger l/a ratio.

A straightforward explanation of why the grating structure becomes locally reacting for $l/a\gg 1$ is that the interaction of the fields between adjacent slots of a grating is reduced as the slot depth is much larger than the slot width, and therefore each slot's response to an incident field is basically isolated (locally reacting). It then follows that since the acoustic leakage of the partition walls actually increases the interactions between adjacent slots, an impedance grating with a finite Z will be less locally reacting compared to a

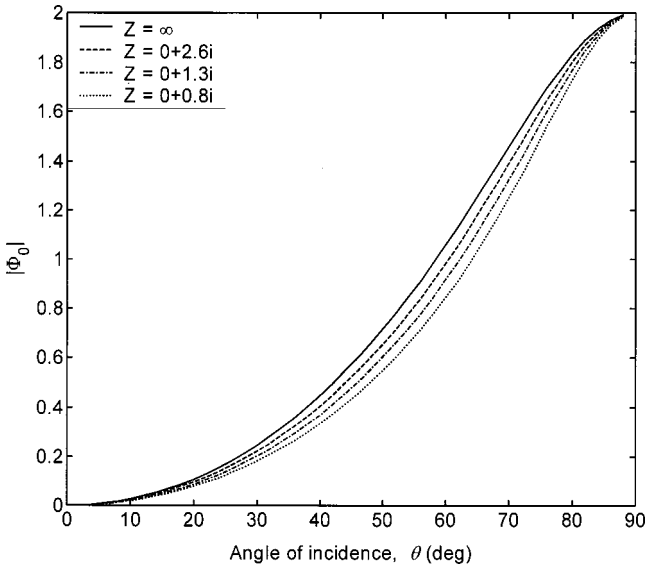


FIG. 5. The magnitude of the mode amplitude $|\Phi_0|$ versus angle of incidence (same parameters as in Fig. 2).

TABLE I. Amplitudes of the surface mode (one diffraction wave).

	$Z=\infty,\ kl=0.57,\ \theta=34°$			
	$\Phi_m \exp(i\beta_m l)$		$\Phi_{-m} \exp(i\beta_{-m} l)$	
	Real	Imaginary	Real	Imaginary
$m=1$	1.01×10^{-2}	1.64×10^{-3}	-1.11×10^{-2}	-1.80×10^{-3}
2	3.41×10^{-3}	5.51×10^{-4}	-3.58×10^{-3}	-5.79×10^{-4}
3	1.71×10^{-3}	2.76×10^{-4}	-1.77×10^{-3}	-2.86×10^{-4}
4	1.02×10^{-3}	1.64×10^{-4}	-1.05×10^{-3}	-1.69×10^{-4}
5	6.65×10^{-4}	1.07×10^{-4}	-6.81×10^{-4}	-1.10×10^{-4}
6	4.62×10^{-4}	7.46×10^{-5}	-4.71×10^{-4}	-7.61×10^{-5}
7	3.34×10^{-4}	5.40×10^{-5}	-3.40×10^{-4}	-5.50×10^{-5}
8	2.49×10^{-4}	4.02×10^{-5}	-2.53×10^{-4}	-4.09×10^{-5}
9	1.90×10^{-4}	3.07×10^{-5}	-1.93×10^{-4}	-3.12×10^{-5}
10	1.47×10^{-4}	2.38×10^{-5}	-1.49×10^{-4}	-2.41×10^{-5}

rigid grating with the same l/a ratio (as can be seen in Fig. 2).

Figure 5 shows the variation of mode amplitude Φ_0 with angle of incidence for four different values of the leakage impedance, using the same structural parameters as in Fig. 2. Smaller reactances (more leakage) lead to lower-mode amplitudes, but the impedance has no effect on the normal and grazing incidence. For normal incidence, due to the symmetry, there is no transverse velocity across the leaky partition walls so they behave as rigid walls. For grazing incidence, we obtain a result analogous to the Lloyd's mirror effect in optics, which requires Φ_0 to be -2 for all values of wall impedance.

The amplitudes of the surface modes of $\pm m$th order, for $m=1,...,10$, are given in Table I for a grating with $a=0.0256$ m, $l=0.015$ m, and $Z=\infty$. In the table, the exponential factor is introduced because we intend to calculate the fields of the surface modes in Eq. (29) starting at $z=l$ and going above, so $\Phi_m e^{i\beta_m z}=(\Phi_m e^{i\beta_m l})e^{i\beta_m(z-l)}$. As seen, the amplitudes $\Phi_m \exp(i\beta_m l)$ and $\Phi_{-m} \exp(i\beta_{-m} l)$ have nearly the same magnitudes (real and imaginary) but opposite signs, and they are much smaller than unity, the incident wave amplitude. This has been observed for different angles of incidence and frequencies that satisfy the inequality of Eq. (45). Therefore, each pair of such modes $(\pm m)$ will form a small standing wave immediately above the grating. The total field of all surface modes does not show a net energy flow.

The energy conservation condition was examined for the numerical results presented in this section. Now, the fact that only the reflection wave exists in the scattered far field, Eq. (43) says that the incident energy is equal to the reflection energy, or $|1+\Phi_0|^2=1$. This has been tested and is satisfied to within $\pm10^{-12}$ for the rigid and pure reactance gratings.

B. Two diffraction modes: The reflection mode plus the $m=-1$ order diffraction mode

An interesting case of grating scattering for homogeneous plane-wave incidence arises when there exists one diffraction mode in addition to the reflection wave. When the angle of incidence and frequency satisfy

$$1+\sin\theta>\lambda/a>\max[1-\sin\theta,\ (1+\sin\theta)/2], \tag{50}$$

then the $m=-1$ order mode is a diffraction mode. The far field can be expressed as

$$\phi_{\text{far}}(x,z)=\exp(i\xi_0 x-i\eta_0 z)+R_0\exp(i\xi_0 x+i\eta_0 z)+\Phi_{-1}\exp(i\alpha_{-1}x+i\beta_{-1}z), \tag{51}$$

where $R_0=1+\Phi_0$ and Φ_{-1} are the reflection and diffraction coefficients, respectively. Figure 6 shows the energy-efficiency plots for these two modes as a function of kl, for $\lambda/a=1.2$, $\theta=37.6°$, and three different values of Z. For the rigid grating, as seen from Fig. 6(a), the two modes alternatively dominate the scattered field as the grating height varies. Their energy efficiencies demonstrate the periodicity over the range of the figure with a period of about 16. The Brewster-angle anomaly[31] is observed at $kl\approx7.5$ and 23.5. At these points, the specular reflection is extinguished and all the energy is converted into the diffraction mode in the backscattering direction. A very similar result was obtained by DeSanto [Fig. 5(a) of Ref. 22], who solved the scattering problem of a rigid, comb-like grating by using the Bloch–Floquet expansion and a meromorphic function analysis. When the wall impedance has finite reactance, Figs. 6(b) and (c), the periodicity in kl of the energy efficiencies becomes large. The Brewster-angle anomaly first occurs at $kl\approx12$ for $Z=4i$ and at $kl\approx16$ for $Z=2i$, respectively.

The amplitudes of the surface modes for the $\pm m$th order, $m=1,...,10$, are given in Table II, using the same parameters as above (Fig. 6) for the $Z=\infty$ case. It is of interest to note that since the $m=-1$ order becomes the diffraction mode, it is now the pair of $\Phi_m \exp(i\beta_m l)$ and $\Phi_{-m-1} \exp(i\beta_{-m-1} l)$ that has nearly the same complex values but opposite signs. As a consequence, the total field due to the surface modes does not show significant net energy flow either, as in the preceding subsection. Finally, the energy conservation condition (43) has been tested for this case and is satisfied with an accuracy of about $\sim10^{-3}$, for the rigid and pure reactance gratings.

V. SURFACE-WAVE DISPERSION ($ka\ll1$)

For a comb-like grating structure with very small ka, a simple analysis (for example, in Ref. 13) gives the following dispersion relation:

$$\xi_0=k\sqrt{1+\tan^2(kl)}. \tag{52}$$

This approximate dispersion equation may also be obtained from the boundary condition given for the equivalent impedance surface defined by Eq. (49). It has a banded-frequency structure, i.e., passbands: $n\pi>kl>(n+1/2)\pi$ and stopbands: $(n+1/2)\pi>kl>(n+1)\pi$, which are equal width. However, this approach does not account for the effects of the grating period.

According to Tolstoy's explanation[13] of the surface-wave formation mechanism from the energy-exchange viewpoint, the existence of surface waves guided by a periodic corrugated surface corresponds to an eigenvalue problem that has nontrivial solutions with the absence of incident waves. Just as the natural frequencies of a vibrational system can be determined by neglecting the excitation forces in the system equations, the dispersion relation for surface waves

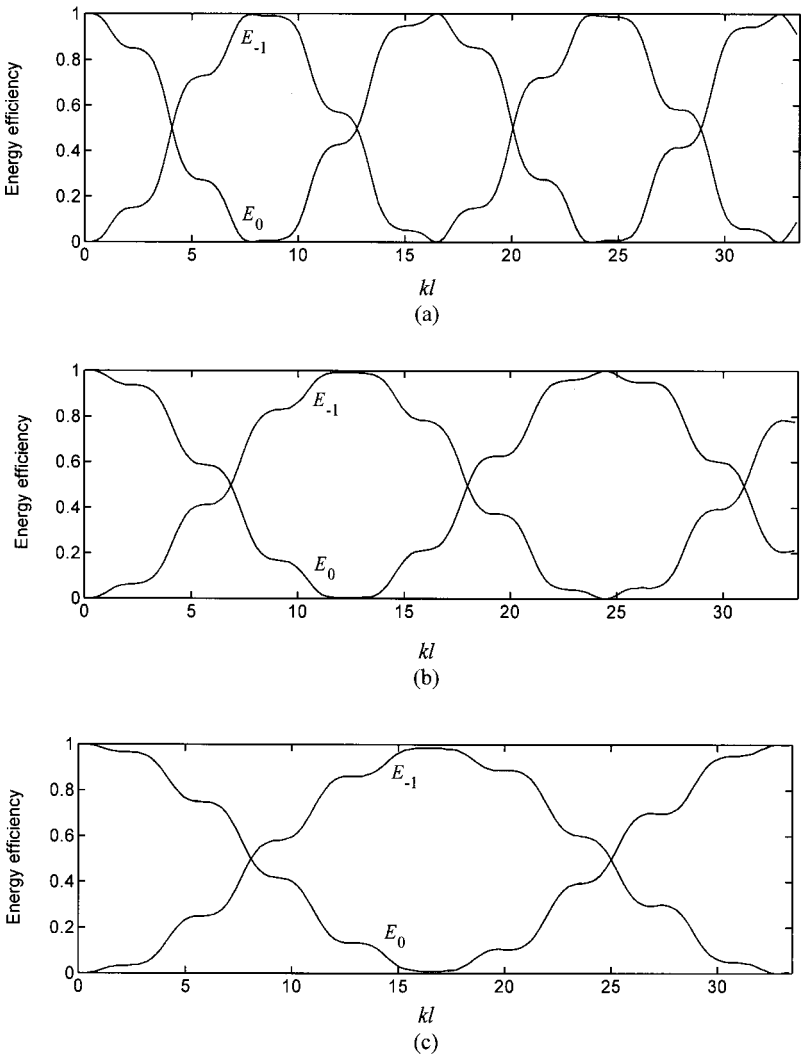


FIG. 6. The reflection and diffraction energy efficiencies versus kl: (a) $Z=\infty$ (rigid grating); (b) $Z=4.0i$; and (c) $Z=2.0i$ ($\lambda/a=1.2, \theta=37.6°$).

over a periodic surface structure is an intrinsic property of the air–structure system and is independent of the incident field. Thus, Eq. (35) or (38) can be used to predict the phase velocities of surface waves over an impedance grating by neglecting the right-hand-side term.

For each kl, the term ξ_0 was systematically varied to determine the roots of

$$\det[(a_{jk}+\pi/4\delta_{jk})]=0, \tag{53}$$

TABLE II. Amplitudes of the surface mode (two diffraction waves).

	$Z=\infty,\ kl=11.0,\ \theta=37.6°$			
	$\Phi_m \exp(i\beta_m l)$		$\Phi_{-m} \exp(i\beta_{-m} l)$	
	Real	Imaginary	Real	Imaginary
$m=1$	5.23×10^{-2}	-5.35×10^{-2}	0.691	0.319
2	1.51×10^{-2}	-1.56×10^{-2}	-5.39×10^{-2}	5.51×10^{-2}
3	7.09×10^{-3}	-7.39×10^{-3}	-1.53×10^{-2}	1.59×10^{-2}
4	4.06×10^{-3}	-4.24×10^{-3}	-7.17×10^{-3}	7.47×10^{-3}
5	2.58×10^{-3}	-2.69×10^{-3}	-4.09×10^{-3}	4.27×10^{-3}
6	1.74×10^{-3}	-1.82×10^{-3}	-2.60×10^{-3}	2.71×10^{-3}
7	1.23×10^{-3}	-1.29×10^{-3}	-1.76×10^{-3}	1.83×10^{-3}
8	8.94×10^{-4}	-9.35×10^{-4}	-1.24×10^{-3}	1.29×10^{-3}
9	6.64×10^{-4}	-6.94×10^{-4}	-8.99×10^{-4}	9.40×10^{-4}
10	5.00×10^{-4}	-5.23×10^{-4}	-6.67×10^{-4}	6.97×10^{-4}

thus obtaining the dispersion relation $\xi_0(k)$. We start by assuming rigid partition walls ($Z=\infty$). A sequence of real roots ξ_0' is obtained, each root separated from the next by $2\pi/a$. This is the same spacing as that between successive $\alpha_m=\xi_0+2\pi m/a$. The sequence of roots ξ_0' has a one-to-one correspondence with the set of wave numbers $\{\alpha_m, m=0, \pm1, \pm\ldots\}$. For this case of small ka, we select as ξ_0 the root ξ_0' that corresponds to the limiting ($a\to0$) curve of Eq. (52). Taking the case of $l/a=12$, the dispersion relation shown in Fig. 7 is obtained. The surface-wave dispersion has a passband/stop-band structure. For the nth passband, $n\pi < kl < (n+1/2)\pi$. The dispersion curve does not quite fill all of the passband except in the limit of $a\to0$. All curves are above the straight dashed line $\xi_0 l=kl$, and therefore the surface waves propagate slower than sound in free air. Since, from Eq. (5) we have $\sin\theta>1$, this surface wave cannot be excited by a homogeneous incident plane wave.

The simple approach in Eq. (52) can be improved to include the effect of the grating period, introducing a modified formula[15,3]

$$\xi_0=k\sqrt{1+\tan^2\left[kl\left(1-\frac{a\ln 2}{\pi l}\right)\right]}. \tag{54}$$

Figure 8 shows the comparison of the exact results to the

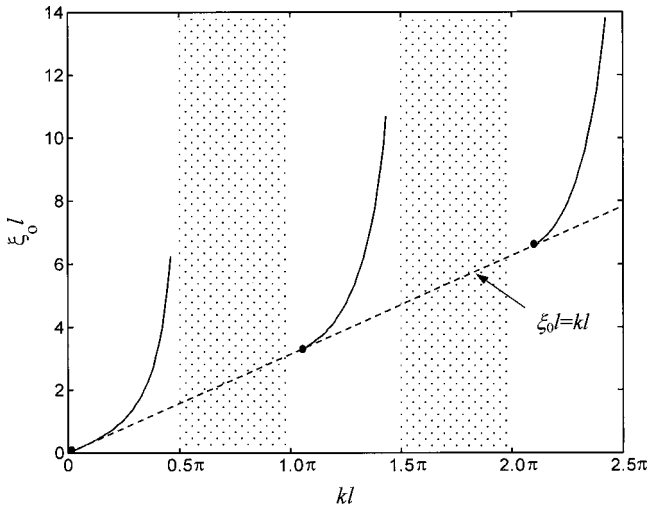


FIG. 7. Predicted dispersion curves (solid) of surface waves above a rigid grating surface ($Z=\infty$, $l/a=12$), showing the passband and stop-band structure (the circles indicate the starting points of the curves).

approximate formulas (52) and (54) for the dispersion curve in the base passband. In the case of a rigid grating, the approximate formulas show good agreement with the exact result in the lower-frequency portion of the band, but, in the higher-frequency portion, deviations are apparent. While Eq. (54) is as simple as Eq. (52), it gives better agreement than the latter since it includes the effect of grating period. But, both formulas overestimate the dispersion of the surface waves in that portion. Figure 8 also shows that the impedance of the partition walls has an important effect on the dispersion. Compared to the rigid grating case, the inductive reactance results in less dispersion. Adding resistance tends to offset the effect of the reactance. Therefore, the inductive reactance in the impedance grating could help in reducing the propagation distortion of signals carried by surface waves over the grating.

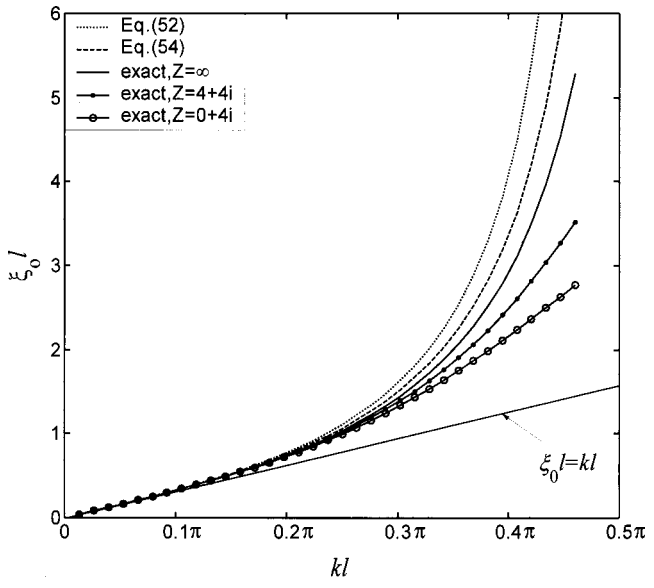


FIG. 8. Comparison of the dispersion curves obtained from Eqs. (52) and (54), and from the exact solution for three different impedance values ($l/a=4$).

VI. SURFACE WAVE DOMINATED BY A HIGH-ORDER SURFACE MODE

At low frequencies ($ka \ll 1$), homogeneous plane waves cannot excite effective surface waves over a grating structure, unless there are some defects in the grating (e.g., discontinuities) which can cause energy redistribution among the modes. We now consider high frequencies for which the incident wavelength is comparable to the grating period. In the following, we will show through numerical examples that under certain conditions, related to the resonance anomalies,[32] a specific surface mode of higher order ($|m| > 0$) can be strongly excited and net surface-wave propagation can be obtained.

A. The resonance of a surface mode under homogeneous plane-wave incidence

Consider homogeneous plane-wave incidence with a wavelength λ that is comparable to the grating period a. As $\sin\theta = \xi_0/k$, the normalized wave number in the x direction of the incident wave, approaches a pole of the amplitude coefficient of some surface mode [see Eq. (30)], this mode generally tends to become dominant and a net surface wave is expected to propagate along the grating surface. The exception is when there is a zero near the pole, offsetting the effect of the pole. First, consider the case in Sec. IV A, where the only diffraction mode is the reflection wave, i.e., the zero-order mode, and the condition (45) holds. From Eqs. (30) and (25), we see that as $\sin\theta$ approaches $\lambda/a - 1$, $\beta_{-1} \to 0$, and the amplitude Φ_{-1} of the $m=-1$ order mode becomes very large. Here, for $\sin\theta$ to approach $\lambda/a - 1$, we need to have $1 < \lambda/a < 2$ (only the case of $\theta > 0$ is considered since the problem is symmetrical about $\theta = 0$). For example, choosing $\lambda/a = 1.92$, we find that at $\theta_p = \arcsin(0.92) \approx 67°$, $\beta_{-1} = 0$, and therefore, $\sin\theta_p$ corresponds to a real pole of Φ_{-1}. To compare this resonant surface mode with the reflection wave and to see the near-field distribution, we calculate the vertical profile of the scattered sound field above the grating [the expansion Eq. (29) is truncated at $m = \pm 15$ in the actual calculation]. Figure 9 shows the scattered field magnitude as a function of height ($z/l \geqslant 1$) for several angles of incidence (the incident wave, and therefore the reflection wave, has unit amplitude). As can be seen, the scattered field magnitude shows basically a bulk-wave feature at a small angle of incidence ($\theta = 16°$). The surface wave field (the $m=-1$ order) then grows as θ increases, and finally the surface wave exceeds the reflection wave at $\theta = 64°$. Figure 10 gives the amplitude variation of the first eight surface modes versus angle of incidence. The $m=-1$ order surface mode becomes resonance as $\sin\theta$ approaches $\sin\theta_p$, while the amplitudes of the other surface modes remain relatively constant. It is also noted that when $\sin\theta$ is very close to $\sin\theta_p$, $1/\beta_{-1} \to \infty$, or the vertical extent of the surface wave becomes very large. Eventually, when the inequality Eq. (45) becomes equality, the Rayleigh wavelength is thus obtained for which the $m=-1$ order surface mode just turns to the $m=-1$ order diffraction mode at grazing angle; this corresponds to the Rayleigh anomaly.[31]

The integration in Eq. (30), which depends on the grating depth l and the function $f(t)$, contributes to the pole/zero

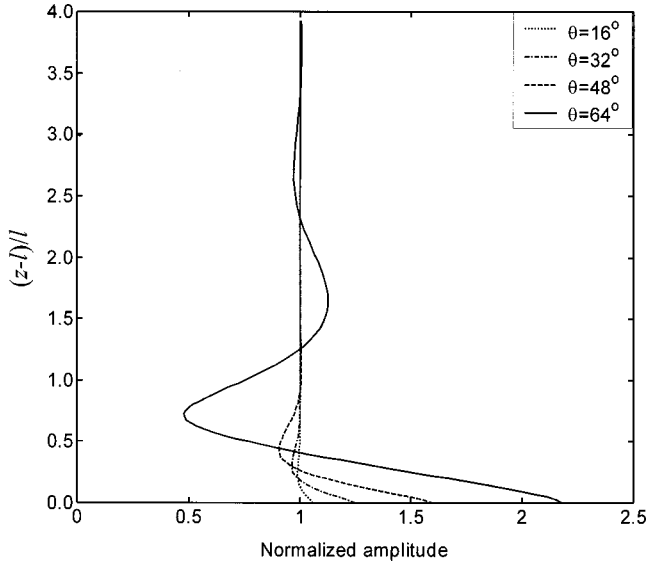


FIG. 9. Vertical profile of the magnitude of the scattered sound field (the reflection wave plus space harmonics) above a rigid grating. ($Z=\infty$, $a=0.06$ m, $l=0.128$ m, $f=2.95$ kHz.)

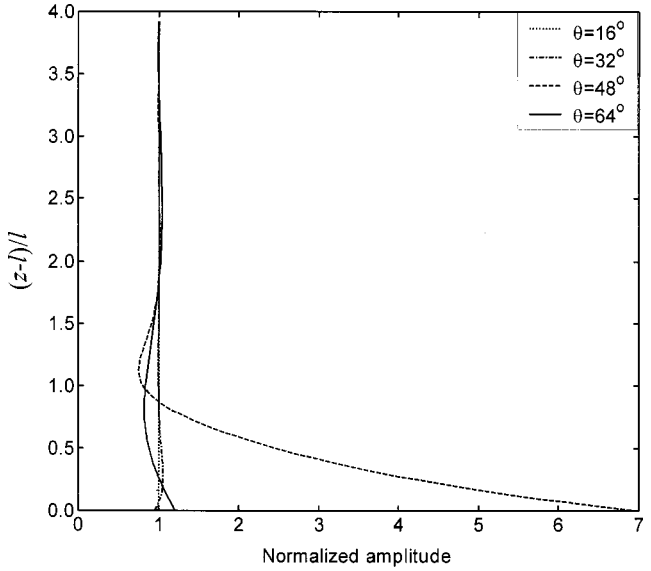


FIG. 11. Vertical profile of the magnitude of the scattered sound field (the reflection wave plus space harmonics) above a rigid grating. ($Z=\infty$, $a=0.06$ m, $l=0.08$ m, $f=2.95$ kHz.)

distribution of the mode amplitude Φ_m. Consider a second example for a smaller grating depth. Figures 11 and 12, respectively, show the vertical profile of the magnitude of the scattered sound field and each mode's amplitude versus angle of incidence. As seen from the figures, the surface wave is most strongly excited at $\theta=48°$ and is dominated by the $m=-1$ order mode. Close to the surface, the amplitude of this mode is five times higher than that of the reflection wave. Obviously from Fig. 12, the resonance position $\sin\theta_p$ is shifted from the sin(67°) to sin(48°), though the ratio λ/a has been kept the same. This means that changes in the grating depth l can cause, through the integration in Eq. (30), relocations of the poles and zeros [in this case, a zero is moved near the sin(67°) and a new pole position is located to sin(48°)]. This new pole is actually a complex number, since the mode amplitude does not grow to infinity as in the case of the real pole in the first example when $\sin\theta \rightarrow \sin\theta_p$. To explain this, let $\zeta=\sin\theta$ and consider the complex mode amplitude function $\Phi_{-1}=\Phi_{-1}(\zeta)$. With the aid of the analytical continuation, $\Phi_{-1}(\zeta)$ can be defined in the complex plane ζ. Then, some poles and zeros of $\Phi_{-1}(\zeta)$ arise as complex numbers. To precisely locate these poles and zeros, the integration in Eq. (30) needs to be completed. As the real-valued $\sin\theta$ approaches the $\mathrm{Re}(\zeta_p)$ of any complex pole ζ_p, Φ_{-1} will become pronounced (not infinity). This corresponds to a resonance anomaly,[32] where it is now a surface mode instead of a diffraction mode that becomes resonance.

The influence of the leakage impedance of the grating of the resonance of the surface mode (Φ_{-1}) is given in Fig. 13,

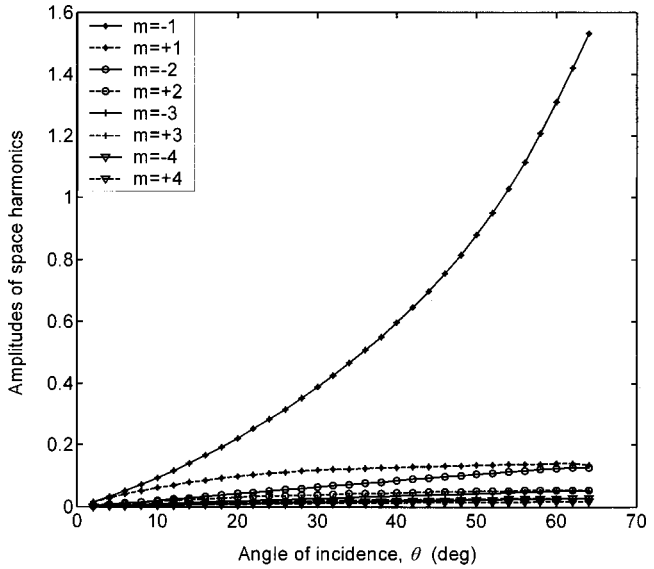


FIG. 10. The amplitudes $|\Phi_m \exp(i\beta_m l)|$ of the first eight space harmonics, showing the $m=-1$ order surface mode becomes pronounced as θ increases. ($Z=\infty$, $a=0.06$ m, $l=0.128$ m, $f=2.95$ kHz.)

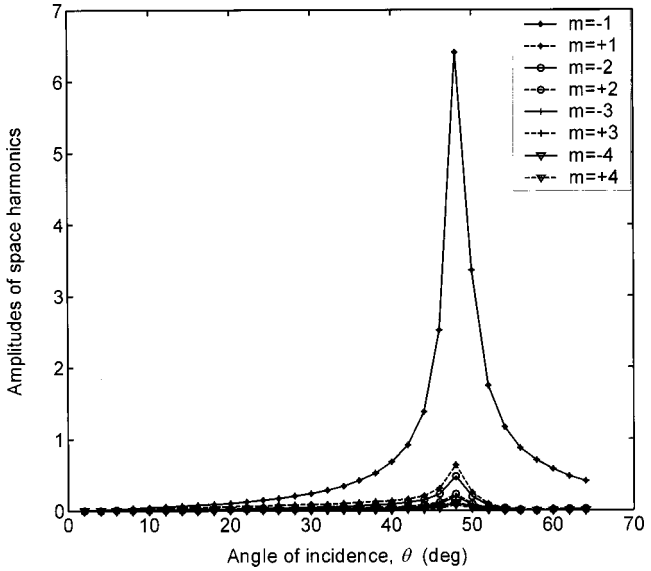


FIG. 12. The amplitudes $|\Phi_m \exp(i\beta_m l)|$ of the first eight space harmonics, showing the $m=-1$ order surface mode becomes resonance at $\theta=48°$. ($Z=\infty$, $a=0.06$ m, $l=0.08$ m, $f=2.95$ kHz.)

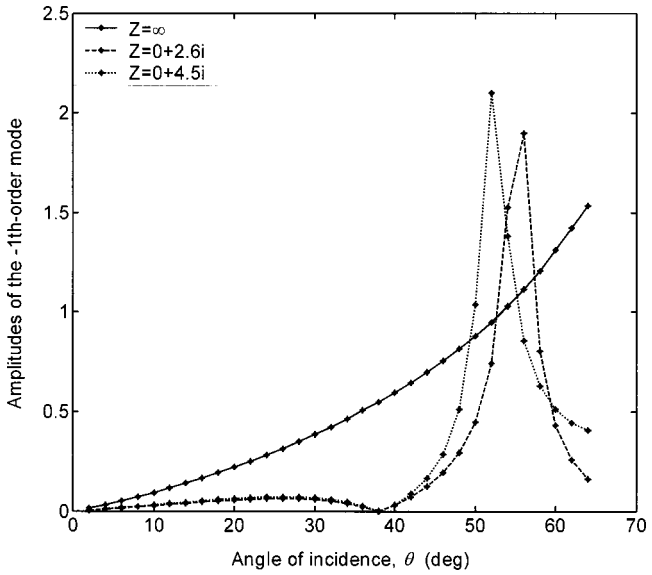


FIG. 13. The amplitude $|\Phi_{-1}\exp(i\beta_{-1}l)|$ of the $m=-1$ order surface mode versus θ, showing the impedance effect on the resonant position. ($a=0.06$ m, $l=0.128$ m, $f=2.95$ kHz, the solid line from Fig. 10.)

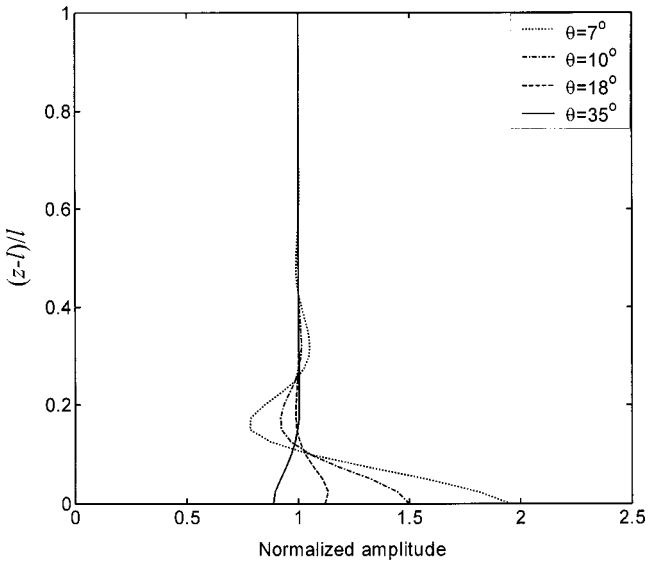


FIG. 14. Vertical profile of the magnitude of the incident wave plus space harmonics above a rigid grating. ($Z=\infty$, $a=0.1$ m, $l=0.287$ m, $f=3.66$ kHz.)

using the same parameters as in the first example. Compared to the rigid grating ($Z=\infty$), the resonant position is shifted to smaller angles of incidence with the introduction of an inductive reactance.

The $m=-1$ order resonant mode Φ_{-1} discussed above, combined with all other nonresonant surface modes, results in a net backscattering surface wave [since $\alpha_1<0$ as seen from Eq. (25)]. A net forward-scattering surface wave dominated by, for example, the $m=1$ order surface mode, can also be generated with homogeneous plane-wave incidence. The third example to be considered is the case in Sec. IV B, where two diffraction modes, i.e., the reflection wave and the $m=-1$ order diffraction mode, exist and the condition (50) holds. By choosing appropriate λ and l, tuning θ, we can excite resonance behavior of the $m=1$ order mode. For instance, let $\lambda/a=0.93$ and $l/a=2.87$; by adjusting the angle of incidence, we find the resonant position for the mode Φ_1 at $\theta=7°$ (see Fig. 14). The figure shows the vertical profile of the total field of the first 28 surface modes ($m=1-14$ and $-2--15$) plus the incident wave. The reason why we included the incidence wave (unit amplitude) in the comparison of vertical profiles instead of the reflection wave, as we did for Figs. 9 and 11, is that the latter is now no longer constant as angle of incidence varies and is therefore not suitable as a reference. The amplitude variation of each mode with angle of incidence is plotted in Fig. 15. The $m=1$ order surface mode becomes pronounced as $\sin\theta$ approaches $\sin(7°)$, while the amplitudes of other surface modes remain small. This results in a net forward-scattering surface wave over the grating. A similar procedure can be applied for tuning other resonant surface modes.

It is of interest to observe the kl values for these three examples. They are $kl=2.22\pi$ for the first example, 1.39π for the second, and 6.17π for the third, which fall within the third, the second, and the sixth passbands, respectively. However, their phase velocities cannot be obtained directly from the curves in Fig. 7, because in obtaining the curves we assumed no diffraction modes in the expansion of Eq. (29). Instead, the phase velocity of the surface wave in each example is directly calculated from the values of α_m and k that correspond to the resonant position of the surface mode.

B. Locating the resonance of a surface mode in the $k-\alpha$ plane and its dispersion curve for $ka\approx 1$

For a given grating (a and l), the resonance of surface modes depends on both the frequency and the angle of the incidence. Locating the pair (f, θ) corresponding to the resonance of the mth-order mode can be nontrivial because the positions of the complex poles/zeros after analytical continuation of the amplitude Φ_m are unknown (except for the zeros of β_m). However, the banded structure shown in Fig. 7 for

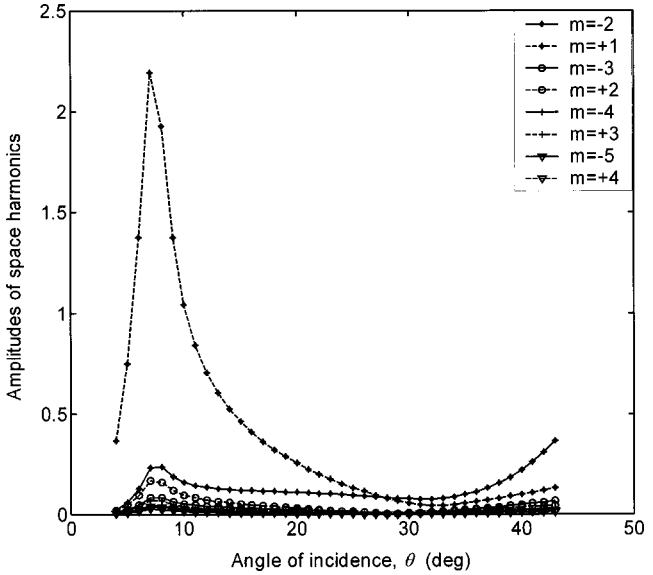


FIG. 15. The amplitudes $|\Phi_m\exp(i\beta_m l)|$ of the first eight space harmonics, showing the $m=1$ order surface mode (forward scattered) becomes resonance at $\theta=7°$. ($Z=\infty$, $a=0.1$ m, $l=0.287$ m, $f=3.66$ kHz.)

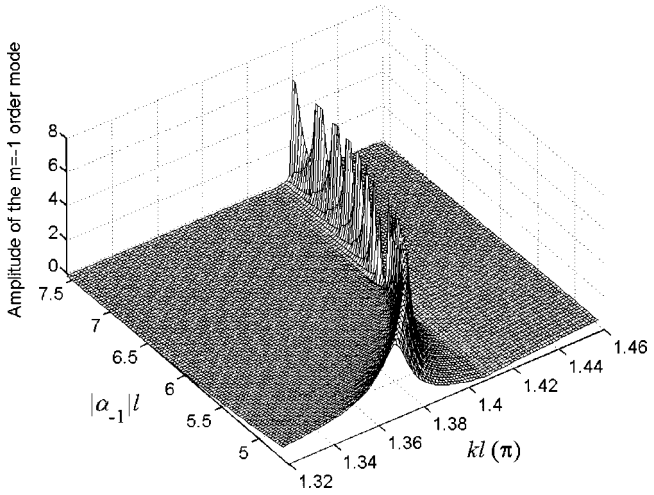


FIG. 16. The resonant position [$\Phi_{-1}\exp(i\beta_{-1}l)$ is truncated at 6] of the $m=-1$ order surface mode in the $kl-|\alpha_{-1}|l$ plane.

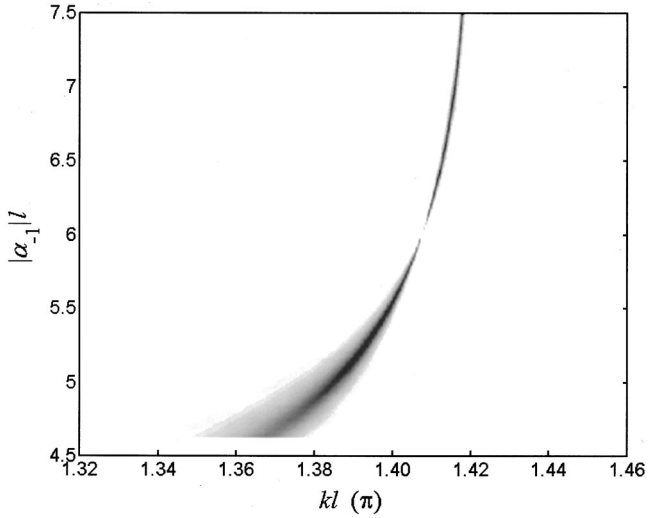


FIG. 17. The dispersion curve for the $m=-1$ order resonant surface mode obtained from a top view of the amplitude surface in Fig. 16 but with a refined mesh.

surface-wave dispersion can be consulted in determining the possible existence region of the resonance of a high-order surface mode. By scanning both frequency and angle in the passband regions in Fig. 7, and observing the maxima of surface mode amplitudes, we can locate the resonant positions or the values of (f, θ) numerically. Additional constraints may apply to the possible existence region of resonance, such as the inequality Eq. (45) or Eq. (50), under which certain diffraction modes exist as well. In the following, instead of the pair (f, θ) we consider alternatively a dimensionless pair (kl,$|\alpha_m|l$) by using the variable transforms $k=2\pi f/c$ and $\alpha_m=k\sin\theta+2\pi m/a$, and attempt to locate the resonant positions in the $kl-|\alpha_m|l$ plane.

Again, consider the $m=-1$ order surface mode in the second example of the preceding subsection, i.e., the rigid grating case with $a=0.06$ m and $l=0.08$ m. We limit our searching scope within the second passband of Fig. 7 with $|\alpha_1|l>kl$ (since the surface wave velocity is lower than the sound speed in free air). Further limits are $|\alpha_1|l=|2\pi l/a-kl\sin\theta|>|2\pi l/a-kl|$ and $2\pi l/a>|\alpha_{-1}|l$ since $\theta\geqslant 0$. Then, we select a scanning region of $kl\in[1.32\pi,1.46\pi]$ and $|\alpha_{-1}|l\in[4.5,7.6]$, which corresponds to $f\in[2800$ Hz,3100 Hz] and $\theta\in[10°,65°]$, respectively. Within this region, the amplitude response $|\Phi_{-1}\exp(i\beta_{-1}l)|$ of the $m=-1$ order surface mode is calculated at discrete values of (kl,$|\alpha_{-1}|l$) and is plotted as a function of the two variables in Fig. 16 in a 3D format. For the large values of $|\alpha_{-1}|l$ (>6, for the geometry), a detailed numerical search with very fine mesh for the amplitude was conducted. The search revealed that the amplitude increased to very large values but varied continuously with the occasional extremely large peak. For display purposes, we have truncated the amplitude at $|\Phi_{-1}\exp(i\beta_{-1}l)|=6$ in the figure with reduced mesh density. From Fig. 16, one is able to find not only the resonant position in terms of (f, θ) and the order of magnitude of the surface mode (bearing in mind that the incident wave amplitude has unit amplitude), but also the dispersion relation of it. Note that each surface mode has to obey its own dispersion relation in order to become resonant and to propagate along the grating surface; the locus of the resonant positions, therefore, actually constitutes the dispersion curve of this surface mode. Figure 17 shows the dispersion curve thus formed by simply taking a top view of the amplitude surface in Fig. 16 but with a refined mesh. The dispersion relations for $m=1$ order and other higher-order surface modes can be obtained in the same way.

VII. CONCLUDING REMARKS

An exact analytical solution has been obtained for the plane-wave scattering by a comb-like, impedance grating. The predictions for the Brewster-angle anomaly for a rigid grating are consistent with a solution[22] obtained using a meromorphic analysis. The advantages of this approach over modal model and meromorphic analysis methods are that the coefficients of the Bloch–Floquet expansion of the scattered field, either above the grating or inside the slots of the grating, are explicitly specified by the tangent velocity difference across a partition wall, and the resulting system of linear equations is numerically stable. Therefore, if it is possible to measure the tangent velocity difference across a partition wall experimentally, then one can directly calculate all the mode amplitudes (diffraction and surface modes) with these explicit expressions without solving the integral equation. Furthermore, based on these explicit expressions, many properties of the scattering process, such as the asymptotic behavior of the expansions and the poles of the diffraction amplitudes, are easier to access. The drawback is it is only suitable for comb-like gratings.

The equivalence of the impedance grating to a flat impedance surface has been demonstrated at low frequencies, where only the specular reflection wave exists in the scattered far fields. To achieve the field-independent or locally reacting equivalent surface impedance, it is necessary that $l/a\gg 1$ and $\lambda/a\gg 1$. The acoustic leakage through the partition walls of a grating contributes to the nonlocal reaction of the equivalent impedance.

Simple, approximate formulas provide a relatively accurate estimation of the dispersion relation in the low fre-

quency range of the base passband. An inductive component in the leakage impedance can reduce the dispersion of surface waves in the high-frequency range of the same band. The numerical results at low frequencies also show that homogeneous plane-wave incidence cannot excite effective surface waves on perfect gratings. At high frequencies (wavelengths are comparable to the grating period), it is possible to generate surface waves of large amplitude over a grating in either the backward or forward direction with homogeneous plane-wave incidence. It should be mentioned that with inhomogeneous incident waves or localized sources (e.g., point sources, line sources, and bounded beams), it is possible to generate surface waves over a grating at low frequencies, without necessarily making use of the resonance of high-order surface modes. Nevertheless, the study presented here has demonstrated an interesting technique for surface-wave excitation, making use of periodic structures comparable in size to the source wavelength. This surface mode resonance is the counterpart to the diffraction mode resonance[31,32] well-known in grating theory.

APPENDIX A: THE CONTOUR INTEGRAL

Consider the following inverse Fourier transform:

$$I(x,u)=\frac{1}{4\pi i}\int_{-\infty}^{\infty}F(x,\omega)\frac{e^{i\beta u}}{\beta}d\beta, \tag{A1}$$

where $F(x,\omega)$ is defined by Eqs. (22b) and (18).

When $u>0$, the integral will be convergent in the $\mathrm{Im}(\beta)>0$ half-plane. From Eqs. (18) and (7), the branch points in the integrand of Eq. (A1) are $\beta=\pm k$ (the plane is cut from $-k$ to k), and the poles in the $\mathrm{Im}(\beta)>0$ half-plane are found to be $\beta=\beta_m$, where

$$\beta_0=\eta_0,\quad \beta_m=\begin{cases}\sqrt{k^2-\alpha_m^2}, & k^2>\alpha_m^2\\ i\sqrt{\alpha_m^2-k^2}, & k^2<\alpha_m^2\end{cases},$$

$$\alpha_m=\xi_0+\frac{2\pi m}{a},\quad m=\pm 1,...,\pm\infty, \tag{A2}$$

with β_m being either positive real or positive imaginary. The path of integration in the inverse Fourier transform (A1) can be deformed according to the contour as shown in Fig. 18. Here, we assume η_0 and the real β_m in Eq. (A2) having a small imaginary part, which eventually tends to zero. From

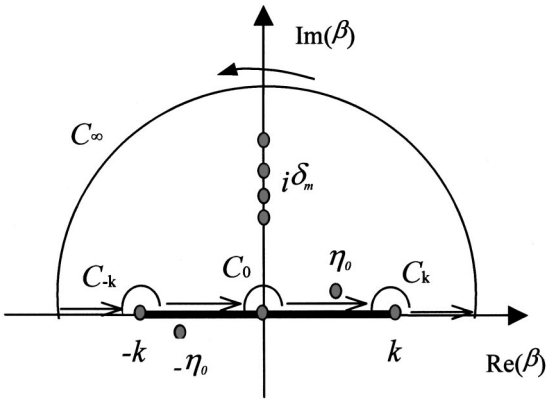


FIG. 18. The contour integral path.

Eqs. (18) and (A2), the residues can be calculated, and the results are

$$\mathrm{Res}\left\{F_1(x,\beta)\frac{e^{i\beta u}}{\beta}\right\}\Big|_{\beta=\beta_m}=\mathrm{Res}\left\{F_2(x,\beta)\frac{e^{i\beta u}}{\beta}\right\}\Big|_{\beta=\beta_m}$$
$$=\frac{\alpha_m}{ia\beta_m^2}e^{ia_m x+i\beta_m u}. \tag{A3}$$

In obtaining Eq. (A3), use has been made of the following identities:

$$\begin{aligned}&\cos[(x+a)\alpha_m]-\cos(x\alpha_m)\exp(-ia\xi_0)\\&\quad=i\sin(a\xi_0)\exp(ix\alpha_m),\\&\cos(x\alpha_m)\exp(ia\xi_0)-\cos[(x-a)\alpha_m]\\&\quad=i\sin(a\xi_0)\exp(ix\alpha_m),\end{aligned} \tag{A4}$$

where α_m is given by Eq. (A2).

Then, the integral in Eq. (A1) can be rewritten as

$$\begin{aligned}\int_{-\infty}^{\infty}[\cdots]d\beta=2\pi i\sum\ \mathrm{Res}[\cdots]-\int_{C_\infty}[\cdots]d\beta\\-\int_{C_{-k}}[\cdots]d\beta-\int_{C_k}[\cdots]d\beta\\-\int_{C_0}[\cdots]d\beta.\end{aligned} \tag{A5}$$

The second term vanishes according to the Jordan's lemma, while it is easy to show the third and fourth terms tend to zero as the circle shrinks to its center. The last one, after some manipulation, gives

$$\int_{C_0}[\cdots]d\beta=-i\pi p(x), \tag{A6}$$

where

$$p(x)=\begin{cases}p_1(x)=\dfrac{\cos[(x+a)k]-\cos(xk)e^{-ia\xi_0}}{\cos(a\xi_0)-\cos(ak)},\\ \qquad na-a/2\leqslant x\leqslant na,\\ p_2(x)=\dfrac{\cos(xk)e^{ia\xi_0}-\cos[(x-a)k]}{\cos(a\xi_0)-\cos(ak)},\\ \qquad na\leqslant x\leqslant na+a/2.\end{cases} \tag{A7}$$

From Eqs. (A1), (A3), (A5), and (A6), we have

$$I(x,u)=2\pi i\left[\sum_{m=-\infty}^{\infty}\frac{\alpha_m}{ia\beta_m^2}e^{i\alpha_m x+i\beta_m u}+\frac{1}{2}p(x)\right], \tag{A8}$$

where $p(x)$ is defined by Eqs. (A7).

When $u<0$, the integral in Eq. (A1) will be convergent in the $\mathrm{Im}(\beta)<0$ half-plane. Using the similar procedure as above but with the poles $\beta=-\beta_m$ located in the $\mathrm{Im}(\beta)<0$ half-plane, we obtain

$$I(x,u)=-2\pi i\left[\sum_{m=-\infty}^{\infty}\frac{\alpha_m}{ia\beta_m^2}e^{ia_m x-i\beta_m u}+\frac{1}{2}p(x)\right]. \tag{A9}$$

Combining Eqs. (A8) and (A9), we have

$$I(x,u)=2\pi i\,\mathrm{sgn}(u)\left[\sum_{n=-\infty}^{\infty}\frac{\alpha_m}{ia\beta_m^2}e^{i\alpha_m x+i\beta_m|u|}+\frac{1}{2}p(x)\right], \tag{A10}$$

where $\mathrm{sgn}(u)$ is the signum function.

Finally, from Eqs. (A1), (A10) and after some algebra, the four inverse Fourier forms in Eq. (23) become

$$\phi_f^{sc}=\begin{cases}\phi_{f1}^{sc}(x,z)=\displaystyle\int_0^l f(t)\left\{\sum_{m=-\infty}^{\infty}\frac{\alpha_m}{ia\beta_m^2}e^{i\alpha_m x+i\beta_m z}[(-2i)\sin(\beta_m t)]\right\}dt, & z\geqslant l,\\ \phi_{f2}^{sc}(x,z)=\displaystyle\int_0^l f(t)\left\{\sum_{m=-\infty}^{\infty}\frac{\alpha_m}{ia\beta_m^2}e^{i\alpha_m x}[\mathrm{sgn}(z-t)e^{i\beta_m|z-t|}-e^{i\beta_m(z+t)}]+\frac{1}{2}[\mathrm{sgn}(z-t)-1]p(x)\right\}dt, & 0\leqslant z\leqslant l,\end{cases} \tag{A11}$$

with $p(x)$ being defined by Eqs. (A7).

APPENDIX B: PROOF OF EQ. (27)

Let

$$\begin{aligned}&\sum_{m=-\infty}^{\infty}\frac{\alpha_m}{(-\beta_m^2)}e^{i\alpha_m x}\\ &=\sum_{m=-\infty}^{\infty}\frac{\left(\xi_0+m\frac{2\pi}{a}\right)}{\left(\xi_0+m\frac{2\pi}{a}\right)^2-k^2}e^{i[\xi_0+m(2\pi/a)]x}\\ &=\sum_{m=-\infty}^{\infty}\tilde{g}\left(m\frac{2\pi}{a}\right),\end{aligned} \tag{B1}$$

where

$$\tilde{g}(u)=\frac{(\xi_0+u)}{(\xi_0+u)^2-k^2}e^{i(\xi_0+u)x}. \tag{B2}$$

From the Poisson's summation formula,[33] we have

$$\sum_{m=-\infty}^{\infty}\tilde{g}\left(\frac{2\pi m}{a}\right)=a\sum_{m=-\infty}^{\infty}g(ma), \tag{B3}$$

where g is the inverse Fourier transform of $\tilde{g}$ and can be expressed as

$$\begin{aligned}g(t)&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{g}(u)e^{itu}\,du\\ &=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{(\xi_0+u)}{(\xi_0+u)^2-k^2}e^{i(\xi_0+u)x+itu}\,du\\ &=\frac{e^{-i\xi_0 t}}{2\pi}\int_{-\infty}^{\infty}\frac{s}{s^2-k^2}e^{is(x+t)}\,ds.\end{aligned} \tag{B4}$$

By using a contour integral and the residue theorem, one finds

$$g(t)=\frac{ie^{-i\xi_0 t}}{2}e^{ik|x+t|}\,\mathrm{sgn}(x+t). \tag{B5}$$

Substituting Eq. (B5) into (B3) results in (note: $|x|\leqslant a/2$)

$$\begin{aligned}\sum_{m=-\infty}^{\infty}\tilde{g}\left(\frac{2\pi m}{a}\right)&=a\sum_{m=-\infty}^{\infty}g(ma)\\ &=\frac{ia}{2}\sum_{m=-\infty}^{\infty}e^{-i\xi_0 ma}e^{ik|x+ma|}\,\mathrm{sgn}(x+ma)\\ &=\frac{ia}{2}\left\{e^{ik|x|}\,\mathrm{sgn}(x)+\sum_{m=1}^{\infty}[e^{-i\xi_0 ma}e^{ik(ma+x)}\right.\\ &\quad\left.-e^{i\xi_0 ma}e^{ik(ma-x)}]\right\}\\ &=\frac{ia}{2}\left\{e^{ik|x|}\,\mathrm{sgn}(x)+e^{ikx}\left[\frac{1}{1-e^{ia(k-\xi_a)}}-1\right]\right.\\ &\quad\left.-e^{-ikx}\left[\frac{1}{1-e^{ia(k+\xi_0)}}-1\right]\right\}\\ &=\frac{ia}{2}\left\{e^{ik|x|}\,\mathrm{sgn}(x)\right.\\ &\quad\left.+i\frac{\sin(kx-a\xi_0)-e^{iak}\sin(kx)}{\cos(ak)-\cos(a\xi_0)}\right\}\\ &=\frac{ia}{2}p(x),\end{aligned} \tag{B6}$$

where $p(x)$ is given by Eq. (A7). Finally, from Eqs. (B6) and (B1), Eq. (27) is proved.

[1] K. M. Ivanov-Shits and F. V. Rozhin, "Investigation of surface waves in air," Sov. Phys. Acoust. **5**, 510–512 (1959).
[2] L. M. Brekhovskikh, "Surface waves in acoustics," Sov. Phys. Acoust. **5**, 3–12 (1959).
[3] R. A. Hurd, "The propagation of an electromagnetic wave along an infinite corrugated surface," Can. J. Phys. **32**, 727–734 (1954).
[4] R. J. Donato, "Model experiments on surface waves," J. Acoust. Soc. Am. **63**, 700–703 (1978).
[5] G. A. Daigle, "Surface waves above porous ground surface," J. Acoust. Soc. Am. Suppl. 1 **85**, S82 (1989).
[6] G. A. Daigle and T. F. W. Embleton, "Surface waves and surface wave devices in atmospheric acoustics," J. Acoust. Soc. Am. Suppl. 1 **88**, S190 (1990).
[7] C. Hutchinson-Howorth and K. Attenborough, "Model experiments on air-coupled surface waves," J. Acoust. Soc. Am. **92**, 2431(A) (1992).
[8] M. R. Stinson, G. A. Daigle, and D. I. Havelock, "The measurement of air-coupled surface waves over a porous surface," in *Proceedings of Inter-Noise 93* (Leuven, Belgium), pp. 1683–1686 (1993).
[9] G. A. Daigle, M. R. Stinson, and D. I. Havelock, "Experiments on surface

waves over a model impedance plane using acoustic pulses,'' J. Acoust. Soc. Am. **99**, 1993–2005 (1996).
[10] M. R. Stinson and G. A. Daigle, ''Surface wave formation at an impedance discontinuity,'' J. Acoust. Soc. Am. **102**, 3269–3275 (1997).
[11] K. B. Rasmussen, ''A note on the calculation of sound propagation over impedance jumps and screens,'' J. Sound Vib. **84**, 598–602 (1982).
[12] G. L. McAninch and M. K. Myers, ''Propagation of quasi-plane waves along an impedance boundary,'' AIAA 26th Aerospace Sciences Meeting, AIAA-88-0179 (1988).
[13] I. Tolstoy, *Wave Propagation* (McGraw-Hill, New York, 1975).
[14] I. Tolstoy, ''The scattering of spherical pulses by slightly rough surfaces,'' J. Acoust. Soc. Am. **66**, 1135–1144 (1979).
[15] L. Kelders, J. F. Allard, and W. Lauriks, ''Ultrasonic surface waves above rectangular-groove gratings,'' J. Acoust. Soc. Am. **103**, 2730–2733 (1998).
[16] W. Lauriks, L. Kelders, and J. F. Allard, ''Surface waves above gratings having a triangular profile,'' Ultrasonics **36**, 865–871 (1998).
[17] J. W. Strutt (Lord Rayleigh), ''On the dynamical theory of gratings,'' Proc. R. Soc. London, Ser. A **79**, 399–416 (1907); or see J. W. Strutt (Lord Rayleigh), *The Theory of Sound* (Dover, New York, 1945), Vol. II, pp. 89–96.
[18] B. A. Lippman, ''Note on the theory of gratings,'' J. Opt. Soc. Am. **43**, 408 (1953).
[19] R. F. Miller, ''On the Rayleigh assumption in scattering by a periodic surface,'' Proc. Cambridge Philos. Soc. **65**, 773–791 (1969).
[20] P. C. Waterman, ''Scattering by periodic surfaces,'' J. Acoust. Soc. Am. **57**, 791–802 (1975).
[21] A. Wirgin, ''Reflection from a corrugated surface,'' J. Acoust. Soc. Am. **68**, 692–699 (1980).
[22] J. DeSanto, ''Scattering from a periodic corrugated structure. II. Thin comb with hard boundary,'' J. Math. Phys. **13**, 336–341 (1972).
[23] J. DeSanto, ''Scattering from a perfectly reflecting arbitrary periodic surface: An exact theory,'' Radio Sci. **16**, 1315–1326 (1981).
[24] R. L. Holford, ''Scattering of sound waves at a periodic, pressure-release surface: An exact solution,'' J. Acoust. Soc. Am. **70**, 1116–1128 (1981).
[25] J. M. Claeys, O. Leroy, and A. Jungman, ''Diffraction of ultrasonic waves from periodically rough liquid–solid surface,'' J. Appl. Phys. **54**, 5657–5662 (1983).
[26] K. E.-A. Van Den Abeele, R. Briers, and O. Leroy, ''Inhomogeneous plane-wave scattering and mode stimulation on periodic rough surfaces,'' J. Acoust. Soc. Am. **99**, 2883–2897 (1996).
[27] F. Odeh and J. B. Keller, ''Partial differential equations with periodic coefficients and Bloch waves in crystals,'' J. Math. Phys. **5**, 1499–1504 (1964).
[28] N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1953).
[29] F. Erdogan, ''Approximation solution of systems of singular integral equations,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **17**, 1041–1059 (1969).
[30] D. S. Jones, *Acoustic and Electromagnetic Waves* (Clarendon, Oxford, 1986).
[31] J. E. Stewart and W. S. Gallaway, ''Diffraction anomalies in grating spectrophotometers,'' Appl. Opt. **1**, 421–429 (1962).
[32] J. R. Andrewartha, J. R. Fox, and I. J. Wilson, ''Resonance anomalies in the lamellar grating,'' Opt. Acta **26**, 69–89 (1979).
[33] A. Papoulis, *The Fourier Integral and its Applications* (McGraw-Hill, New York, 1962).

Acoustic scattering by inhomogeneous spheres

P. A. Martin[a)]
Department of Mathematical and Computer Sciences, Colorado School of Mines, Golden, Colorado 80401-1887

(Received 6 March 2001; revised 19 December 2001; accepted 20 February 2002)

Acoustic scattering problems are considered when the material parameters (density ρ and speed of sound c) are spherically symmetric functions of position. Explicit separated solutions are derived (i) when $\rho(r)=\exp(\beta r)$ and c^{-2} is a linear function of r^{-1}, and (ii) when $\rho(r)=\exp(-\beta r^2)$ and c^{-2} is a linear function of r^2. In both cases, the radial parts of the solutions are given in terms of Coulomb wave functions or Whittaker functions; these are well-studied special functions, closely related to confluent hypergeometric functions. Two problems are discussed in detail: scattering by an inhomogeneous sphere embedded in a homogeneous fluid, and scattering by a homogeneous sphere with a concentric inhomogeneous coating. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1470502]

PACS numbers: 43.20.Fn [ANN]

I. INTRODUCTION

Sound propagation in inhomogeneous media continues to be of interest. Much is known about layered, stratified media, where the properties of the media (density ρ and speed of sound c) depend on one coordinate, z, say, where x, y, and z are Cartesian coordinates (Brekhovskikh, 1960). Such situations have obvious application to underwater acoustics.

For time-harmonic motions of frequency ω the governing equation is

$$\rho \operatorname{div}(\rho^{-1} \operatorname{grad} p) + k^2 p = 0, \tag{1}$$

where $k=\omega/c$ and p is the acoustic pressure. If the density is constant, Eq. (1) reduces to

$$\nabla^2 p + k_0^2 n(\boldsymbol{r}) p = 0, \tag{2}$$

where $\nabla^2 \equiv \operatorname{div}\operatorname{grad}$, $n(\boldsymbol{r})=[c_0/c(\boldsymbol{r})]^2$ is the (square of the) *refractive index* at position $\boldsymbol{r}$, $k_0=\omega/c_0$, and c_0 is a constant sound speed. We assume that $n(\boldsymbol{r})\to 1$ as $r\to\infty$ in all directions; this excludes layered media, of course.

There is a considerable literature on Eq. (2), especially in the context of quantum mechanics; some of this will be mentioned below. In addition, several point-source solutions (Green's functions) are known for various functional forms of $n(\boldsymbol{r})$; see Li *et al.* (1990) for a review.

For scattering problems in acoustics, there are essentially three cases, depending on properties of n. First, suppose that $[1-n(\boldsymbol{r})]$ has compact support, so that $n(\boldsymbol{r}) \equiv 1$ for $r=|\boldsymbol{r}|>a$, say. Suppose further that $n(\boldsymbol{r})$ is smooth for all $\boldsymbol{r}$ in three-dimensional space. Then, one can reduce the scattering problem to an integral equation over D, the finite region in which $n(\boldsymbol{r})\not\equiv 1$. One such is the Lippmann–Schwinger equation (see, for example, Ahner, 1977; Colton and Kress, 1992, Sec. 8.2; and Newton, 1982, Sec. 10.3). Asymptotic approximations are available for $k_0 a \ll 1$ (Ahner, 1977; Kriegsmann *et al.*, 1983); Kriegsmann *et al.* (1983) also discuss Eq. (1) briefly.

Second, we could have situations in which $[1-n(\boldsymbol{r})]$ does *not* have compact support, but is such that $n(\boldsymbol{r})\to 1$ as $r\to\infty$. The corresponding scattering problems are uncommon in acoustics.

Third, we could have $n(\boldsymbol{r})\equiv 1$ outside D with n discontinuous across ∂D, the boundary of D. The corresponding scattering problem will require transmission conditions across the interface ∂D. If the material in D is actually homogeneous, so that $n(\boldsymbol{r})=n_1$, a constant, for all $\boldsymbol{r}\in D$, boundary integral equations over ∂D can be used; see Kleinman and Martin (1988) for a review.

In this paper, we are mainly concerned with this third class of problem: acoustic scattering by a bounded inhomogeneity embedded in an unbounded homogeneous medium. We begin (Sec. II) with a derivation of the partial differential equation (1): this equation governs the acoustic pressure in an otherwise stationary but inhomogeneous compressible fluid. We give this derivation because some textbook discussions are flawed. We then suppose that the inhomogeneity is *spherically symmetric*, so that ρ and c are assumed to be given functions of the spherical polar coordinate r (only). Such problems have been studied by several authors. For problems of the first type [smooth n with $n(r)\equiv 1$ for $r>a$], Ahner (1977) has given low-frequency expansions. For the second type [smooth n, with $n(r)\to 1$ rapidly as $r\to\infty$], Colton (1978) has used so-called "transformation operators," which map solutions of the Helmholtz equation

$$(\nabla^2 + k_0^2) u = 0, \tag{3}$$

into solutions of Eq. (2); see also Colton and Kress (1978). Colton and Kress (1979) have extended this approach to consider related transmission problems; see also Sleeman (1980). Frisk and DeSanto (1970) have exploited the notion of a Jost function from quantum mechanics so as to obtain approximate solutions of Eq. (2).

We consider acoustic scattering by an inhomogeneous sphere of a radius a. The medium in $r>a$ is homogeneous, with density ρ_0 and sound speed c_0. For time-harmonic motions, the acoustic pressure p_0 is governed by Eq. (3). Inside

a)Electronic mail: pamartin@mines.edu

the sphere, $r<a$, the governing equation is Eq. (1). Across $r=a$, we impose continuity of pressure and normal velocity, the latter condition being equivalent to continuity of $\rho^{-1}(\partial p/\partial r)$. This gives a *transmission problem*. If the interior is homogeneous, with $\rho=\rho_1$ and $c=c_1$ constants, the problem can be solved exactly, by separation of variables. We will show that this method can be extended to certain functional forms for $\rho(r)$ and $c(r)$. We also consider scattering by a sphere with a homogeneous core and an inhomogeneous coating.

There are many papers on analogous electromagnetic scattering problems. For an early treatment, see the paper by Wyatt (1962). One popular technique is to replace the inhomogeneous sphere by many concentric layers, and then to use a simple approximation to the refractive index within each layer; see, for example, Perelman (1996). Kai and Massoli (1994) have reported the results of computations with as many as 10 000 layers.

In this paper, we consider the following specific functional forms

(i) $\rho(r)=\rho_1 e^{\beta r}$ and $[k(r)]^2=k_1^2+\alpha r^{-1}$,

(ii) $\rho(r)=\rho_1 e^{-\beta r^2}$ and $[k(r)]^2=k_1^2+\alpha r^2$.

Here, ρ_1, β, k_1^2, and α are adjustable parameters. For both (i) and (ii), explicit solutions of Eq. (1) are derived. The radial parts of these solutions are given in terms of known special functions, namely Coulomb wave functions and Whittaker functions. These solutions then permit the explicit solution of various scattering problems for inhomogeneous spheres. Such solutions can serve as benchmarks for numerical methods, but they also have intrinsic interest.

II. GOVERNING EQUATIONS

In the linear theory of acoustics for an inhomogeneous medium, the basic equation is (Morse and Ingard, 1986, p. 408)

$$\rho\,\mathrm{div}(\rho^{-1}\,\mathrm{grad}\,P)=c^{-2}(\partial^2 P/\partial t^2), \tag{4}$$

where $P(\boldsymbol{r},t)$ is the acoustic pressure at position $\boldsymbol{r}$ and time t, $\rho(\boldsymbol{r})$ is the density, and $c(\boldsymbol{r})$ is the speed of sound. For time-harmonic motions, with $P=\mathrm{Re}\{pe^{-i\omega t}\}$, we obtain Eq. (1).

According to Pierce (1990), Eq. (4) was first given by Bergmann (1946). However, the *derivation* of Eq. (4) does not seem to be well known. In particular, Eq. (4) cannot be derived without mentioning the entropy.

A. Derivation of Bergmann's equation

The exact equations for the motion of an inviscid compressible fluid are (Batchelor, 1967, Sec. 3.6; Ostashev, 1997, Sec. 2.1.1)

$$\frac{\mathrm{D}\tilde{\rho}}{\mathrm{D}t}+\tilde{\rho}\,\mathrm{div}\,\tilde{\boldsymbol{v}}=0, \tag{5}$$

$$\tilde{\rho}\frac{\mathrm{D}\tilde{\boldsymbol{v}}}{\mathrm{D}t}+\mathrm{grad}\,\tilde{P}=\boldsymbol{0}, \tag{6}$$

$$\frac{\mathrm{D}\tilde{S}}{\mathrm{D}t}=0. \tag{7}$$

Here, $\tilde{\rho}$ is the density, $\tilde{\boldsymbol{v}}$ is the velocity, $\tilde{P}$ is the pressure, and $\tilde{S}$ is the entropy per unit mass; all these quantities may depend on $\boldsymbol{r}$ and t. They will be related to the quantities in Eq. (4) later. The material derivative is defined by $(\mathrm{D}f/\mathrm{D}t)(\boldsymbol{r},t)=\partial f/\partial t+(\tilde{\boldsymbol{v}}\cdot\nabla)f$. In writing Eq. (6), we have assumed that there are no body forces. Equation (7) means that the flow is *isentropic* (Batchelor, 1967, p. 156).

We also require an equation of state. As usual, we suppose that

$$\tilde{P}=\tilde{P}(\tilde{\rho},\tilde{S}). \tag{8}$$

It follows that $\mathrm{grad}\,\tilde{P}=\tilde{c}^2\,\mathrm{grad}\,\tilde{\rho}+\tilde{h}\,\mathrm{grad}\,\tilde{S}$, where

$$\tilde{c}^2(\tilde{\rho},\tilde{S})=(\partial\tilde{P}/\partial\tilde{\rho})_{\tilde{S}} \quad\text{and}\quad \tilde{h}(\tilde{\rho},\tilde{S})=(\partial\tilde{P}/\partial\tilde{S})_{\tilde{\rho}}.$$

The temperature $\tilde{T}$ satisfies [Batchelor, 1967, Eq. (3.6.6)]

$$\tilde{\rho}\frac{\mathrm{D}\tilde{T}}{\mathrm{D}t}=\nu\tilde{T}\frac{\mathrm{D}\tilde{P}}{\mathrm{D}t},$$

where ν is the ratio of the specific heat at constant pressure to the coefficient of thermal expansion. Thus

$$\tilde{T}^{-1}\frac{\mathrm{D}\tilde{T}}{\mathrm{D}t}=\frac{\nu\tilde{c}^2}{\tilde{\rho}}\frac{\mathrm{D}\tilde{\rho}}{\mathrm{D}t}=-\nu\tilde{c}^2\,\mathrm{div}\,\tilde{\boldsymbol{v}}, \tag{9}$$

using Eqs. (5), (7), and (8).

For linear acoustics, we suppose that $\tilde{P}=P_0+P_1$, $\tilde{\boldsymbol{v}}=\boldsymbol{v}_0+\boldsymbol{v}_1$, $\tilde{\rho}=\rho_0+\rho_1$, $\tilde{S}=S_0+S_1$, $\tilde{T}=T_0+T_1$, $\tilde{c}=c_0+c_1$, and $\tilde{h}=h_0+h_1$, where the ambient flow is denoted by the subscript 0 and the small acoustic disturbance is denoted by the subscript 1. (The quantities of most interest are P_1, $\boldsymbol{v}_1$, ρ_0, S_0, and c_0.) We require that the ambient flow satisfies Eqs. (5)–(9) exactly, and then we derive a set of linear equations governing the small acoustic disturbance.

From Eq. (8), we obtain $P_0=\tilde{P}(\rho_0,S_0)$ and then

$$P_1=c_0^2\rho_1+h_0S_1, \tag{10}$$

where $c_0^2=\tilde{c}^2(\rho_0,S_0)$ and $h_0=\tilde{h}(\rho_0,S_0)$.

The leading-order equations, governing the ambient flow, follow from Eqs. (5)–(7) and (9)

$$\frac{\partial\rho_0}{\partial t}+\mathrm{div}(\rho_0\boldsymbol{v}_0)=0, \tag{11}$$

$$\rho_0\left\{\frac{\partial\boldsymbol{v}_0}{\partial t}+(\boldsymbol{v}_0\cdot\nabla)\boldsymbol{v}_0\right\}=-\,\mathrm{grad}\,P_0=-c_0^2\,\mathrm{grad}\,\rho_0-h_0\,\mathrm{grad}\,S_0, \tag{12}$$

$$\frac{\partial S_0}{\partial t}+\boldsymbol{v}_0\cdot\mathrm{grad}\,S_0=0, \tag{13}$$

$$\frac{\partial T_0}{\partial t}+(\boldsymbol{v}_0\cdot\nabla)T_0=-\nu c_0^2T_0\,\mathrm{div}\,\boldsymbol{v}_0. \tag{14}$$

Having selected an ambient flow, the acoustic disturbance is then governed by

$$\frac{\partial\rho_1}{\partial t}+\mathrm{div}(\rho_1\boldsymbol{v}_0+\rho_0\boldsymbol{v}_1)=0, \tag{15}$$

$$\rho_0\left\{\frac{\partial \boldsymbol{v}_1}{\partial t}+(\boldsymbol{v}_0\cdot\nabla)\boldsymbol{v}_1+(\boldsymbol{v}_1\cdot\nabla)\boldsymbol{v}_0\right\}$$

$$+\rho_1\left\{\frac{\partial \boldsymbol{v}_0}{\partial t}+(\boldsymbol{v}_0\cdot\nabla)\boldsymbol{v}_0\right\}=-\operatorname{grad}P_1, \tag{16}$$

$$\frac{\partial S_1}{\partial t}+\boldsymbol{v}_0\cdot\operatorname{grad}S_1+\boldsymbol{v}_1\cdot\operatorname{grad}S_0=0. \tag{17}$$

We could derive an equation for T_1, but it will not be needed below.

1. Homogeneous fluid at rest

This is the textbook case (Lighthill, 1978; Pierce, 1989; DeSanto, 1992), where the ambient flow has $\boldsymbol{v}_0=\mathbf{0}$, with P_0, ρ_0, S_0, T_0, c_0, and h_0 all constant. These choices satisfy Eqs. (11)–(14) identically. Then, Eqs. (15)–(17) reduce to

$$\frac{\partial \rho_1}{\partial t}+\rho_0\operatorname{div}\boldsymbol{v}_1=0, \tag{18}$$

$$\rho_0\frac{\partial \boldsymbol{v}_1}{\partial t}=-\operatorname{grad}P_1, \tag{19}$$

$$\frac{\partial S_1}{\partial t}=0, \tag{20}$$

together with Eq. (10). Multiplying Eq. (18) by c_0^2 and Eq. (20) by the thermodynamic coefficient h_0 and adding the results, gives

$$\frac{\partial P_1}{\partial t}+\rho_0 c_0^2\operatorname{div}\boldsymbol{v}_1=0. \tag{21}$$

Then, eliminating $\boldsymbol{v}_1$ between this equation and Eq. (19), we obtain the familiar wave equation, $\nabla^2 P_1=c_0^{-2}(\partial^2 P_1/\partial t^2)$, for the acoustic pressure. The other acoustic quantities, $\boldsymbol{v}_1$, ρ_1, and S_1, can then be calculated in terms of P_1.

In practice, the dependence on entropy is often ignored, so that Eq. (8) is replaced by $\tilde{P}=\tilde{P}(\tilde{\rho})$ when $\tilde{h}=0$ and $S_1=0$. However, we claim that entropy should be retained when the fluid is not homogeneous.

2. Constant entropy

Suppose that S_0 is constant, so that Eq. (13) is satisfied identically. Equation (12) reduces to

$$\frac{\partial \boldsymbol{v}_0}{\partial t}+(\boldsymbol{v}_0\cdot\nabla)\boldsymbol{v}_0=-c_0^2\rho_0^{-1}\operatorname{grad}\rho_0. \tag{22}$$

If we assume that $\boldsymbol{v}_0$ is a constant vector, Eqs. (22) and (11) imply that ρ_0 is constant. Equation (12) then implies that P_0 is a function of t only. But, $P_0=\tilde{P}(\rho_0,S_0)$, so that P_0 is also a constant. Thus, constant S_0 and constant $\boldsymbol{v}_0$ imply that ρ_0 and P_0 are constant too. Also, T_0 must satisfy Eq. (14) with zero on the right-hand side.

Further remarks on the assumption of consant S_0 in the context of stratified media can be found in the book by Ostashev (1997, Sec. 2.2.4).

DeSanto (1992, Appendix 1 A) argued that the second term on the left-hand side of Eq. (22) is negligible, and then deduced that $\boldsymbol{v}_0$ cannot be constant in an inhomogeneous fluid. In fact, if the right-hand side of Eq. (22) does not depend on t, we can integrate, thus showing that $|\boldsymbol{v}_0|$ must grow linearly with t. This is an unpleasant consequence of neglecting entropy.

The derivation given by Morse and Ingard (1986, p. 408) is flawed. They begin (their first displayed equation on p. 408) with

$$0=\frac{\mathrm{D}\rho_0}{\mathrm{D}t}=\frac{\partial \rho_0}{\partial t}+\boldsymbol{v}_0\cdot\operatorname{grad}\rho_0 \tag{23}$$

(in our notation), which is incorrect; cf. Eq. (11). Next, they "add a sound wave, with its velocity **u**, its pressure p, and its additional density change δ," so that $\mathbf{u}\equiv\boldsymbol{v}_1$, $p\equiv P_1$, and $\delta\equiv\rho_1$ in our notation. The following equations are erroneous because, in their Eq. (23), they have **u** in place of $\boldsymbol{v}_0$, so that their **u** is both $\boldsymbol{v}_0$ and $\boldsymbol{v}_1$!

3. Zero ambient velocity

Suppose, instead, that $\boldsymbol{v}_0=\mathbf{0}$. Then, Eqs. (11), (13), and (14) imply that ρ_0, S_0, T_0, c_0, and h_0 do not depend on t. Equation (12) will also be satisfied, provided that

$$c_0^2\operatorname{grad}\rho_0+h_0\operatorname{grad}S_0=\mathbf{0}. \tag{24}$$

This constraint permits us to have spatial variations in c_0^2 and ρ_0 within a stationary fluid.

For the acoustic disturbance, Eqs. (15)–(17) reduce to

$$\frac{\partial \rho_1}{\partial t}+\operatorname{div}(\rho_0\boldsymbol{v}_1)=0, \tag{25}$$

$$\rho_0\frac{\partial \boldsymbol{v}_1}{\partial t}+\operatorname{grad}P_1=\mathbf{0}, \tag{26}$$

$$\frac{\partial S_1}{\partial t}+\boldsymbol{v}_1\cdot\operatorname{grad}S_0=0. \tag{27}$$

Making use of Eq. (10), we combine Eqs. (25) and (27) to give

$$\frac{\partial P_1}{\partial t}+c_0^2\operatorname{div}(\rho_0\boldsymbol{v}_1)+h_0\boldsymbol{v}_1\cdot\operatorname{grad}S_0=0. \tag{28}$$

Using the constraint (24) so as to eliminate S_0 from Eq. (28), we obtain Eq. (21) again (except now ρ_0 and c_0^2 are not required to be constants). Finally, by eliminating $\boldsymbol{v}_1$ between Eqs. (26) and (21), we obtain Bergmann's equation (4), wherein $\rho\equiv\rho_0$, $P\equiv P_1$ and $c\equiv c_0$.

B. Reduced equations

We can reduce Bergmann's equation to an equation without first derivatives by introducing a new dependent variable (Bergmann, 1946); thus, define

$$P=\rho^{1/2}U, \tag{29}$$

whence U is found to satisfy

$$\nabla^2 U+KU=c^{-2}(\partial^2 U/\partial t^2), \tag{30}$$

where

$$K=\tfrac{1}{2}\rho^{-1}\nabla^2\rho-\tfrac{3}{4}\rho^{-2}|\operatorname{grad}\rho|^2 \tag{31}$$

$$=-\rho^{1/2}\nabla^2(\rho^{-1/2}). \tag{32}$$

Equations (29)–(31) [but not Eq. (32)] can be found in Brehovskikh (1960, p. 171).

Evidently, Eq. (30) can be reduced to a partial differential equation with constant coefficients if, for example, $\nabla^2(\rho^{-1/2})=\lambda\rho^{-1/2}$ and $c=c_1$, where λ and c_1 are constants.

For time-harmonic problems, we can write $U=\mathrm{Re}\{ue^{-i\omega t}\}$ and $\boldsymbol{v}=\mathrm{Re}\{\boldsymbol{w}e^{-i\omega t}\}$, whence u satisfies

$$\nabla^2u+(k^2+K)u=0, \tag{33}$$

where $k^2=\omega^2/c^2$, and then the fluid velocity is given by $\boldsymbol{w}=(i\omega\rho)^{-1}\operatorname{grad}(\rho^{1/2}u)$.

In a homogeneous region, we have $\rho=\rho_1$, $c=c_1$, and $k=k_1=\omega/c_1$, all constants. Then, $K\equiv0$ and Eq. (33) reduces to the standard Helmholtz equation. If ρ is constant but c is not, we still have $K\equiv0$ and then Eq. (33) is usually written as Eq. (2); see, for example, Colton and Kress (1992, Chap. 8).

Finally, we can write Eq. (33) as $\nabla^2u+(k_1^2-V)u=0$, where k_1^2 is a constant and $V=k_1^2-k^2-K$, which we recognize as *Schrödinger's equation* with potential V [Newton, 1982, Eq. (10.59)].

III. SPHERICAL SYMMETRY

Introduce spherical polar coordinates, r, θ, and ϕ. Assume that the inhomogeneous medium is spherically symmetric. Then, as $\nabla^2\{f(r)\}=r^{-2}[r^2f']'$ [where $f'(r)\equiv df/dr$], we find that

$$K(r)=r^{-1}(\rho'/\rho)+\tfrac{1}{2}(\rho''/\rho)-\tfrac{3}{4}(\rho'/\rho)^2.$$

Next, we seek solutions of Eq. (33) in the form

$$u(r,\theta,\phi)=u_n(r)Y_n(\theta,\phi), \tag{34}$$

where n is an integer, Y_n is a spherical harmonic, and $u_n(r)$ is to be found by substituting Eq. (34) in Eq. (33). [A typical spherical harmonic is $A_n^mP_n^m(\cos\theta)e^{im\theta}$, where P_n^m is an associated Legendre function and A_n^m is a normalization constant.]

We have

$$\nabla^2(u_nY_n)=u_n\nabla^2Y_n+2(\operatorname{grad}u_n)\cdot(\operatorname{grad}Y_n)+Y_n\nabla^2u_n. \tag{35}$$

But $(\operatorname{grad}u_n)\cdot(\operatorname{grad}Y_n)=0$ because u_n is a function of r and Y_n is a function of θ and ϕ. We also know that r^nY_n is a separated solution of Laplace's equation, so that

$$0=\nabla^2\{r^nY_n\}=r^n\nabla^2Y_n+Y_n\nabla^2\{r^n\},$$

by Eq. (35) and $\nabla^2\{r^n\}=n(n+1)r^{n-2}$ whence $\nabla^2Y_n=-n(n+1)r^{-2}Y_n$, and then Eq. (35) gives

$$\nabla^2(u_nY_n)=\{\nabla^2u_n-n(n+1)r^{-2}u_n\}Y_n.$$

Hence, Eq. (33) reduces to

$$u_n''+2r^{-1}u_n'+[k^2(r)+K(r)-n(n+1)r^{-2}]u_n=0, \tag{36}$$

which is a linear second-order differential equation for $u_n(r)$. If we have solutions of this equation, we can then use the method of separation of variables for various scattering problems involving inhomogeneous spheres. Two such problems are described next.

IV. TWO SCATTERING PROBLEMS

Acoustic scattering by spheres, with various boundary conditions, is a textbook topic (Morse and Ingard, 1986, Sec. 8.2). We shall modify the familiar method of separation of variables so as to treat inhomogeneous spheres.

Consider an inhomogeneous sphere of radius a centered at the origin. Without loss of generality, we can take the incident pressure field as

$$\frac{p_{\mathrm{inc}}}{\rho_0c_0^2}=e^{ik_0z}=\sum_{n=0}^{\infty}(2n+1)i^nj_n(k_0r)P_n(\cos\theta),$$

where $j_n(w)$ is a spherical Bessel function. Then, we can write the total pressure field outside the sphere, in $r>a$, as

$$p_0(r,\theta)=\rho_0c_0^2\sum_{n=0}^{\infty}(2n+1)i^n\{j_n(k_0r)+A_nh_n(k_0r)\}P_n(\cos\theta), \tag{37}$$

where $h_n(w)\equiv h_n^{(1)}(w)$ is a spherical Hankel function and the dimensionless coefficients A_n are to be found. This expression for p_0 satisfies the Helmholtz equation and, moreover, p_0-p_{inc} satisfies the Sommerfield radiation condition at infinity.

A. An inhomogeneous sphere

Inside the sphere ($r<a$), we write

$$u(r,\theta)=\rho_0^{1/2}c_0^2\sum_{n=0}^{\infty}(2n+1)i^nB_nu_n(r)P_n(\cos\theta), \tag{38}$$

where the dimensionless coefficients B_n are to be found and $u_n(r)$ is a solution of Eq. (36) that is regular at $r=0$; some explicit solutions will be given later.

We find A_n and B_n by enforcing the transmission conditions across the interface at $r=a$. Let

$$\rho_a=\lim_{r\to a-}\rho(r)\quad\text{and}\quad\kappa_a=\lim_{r\to a-}[\rho'(r)/\rho(r)],$$

so that ρ_a is the surface value of the interior density. Then, the interface conditions are

$$p_0=\rho_a^{1/2}u\quad\text{and}\quad\rho_0^{-1}\frac{\partial p_0}{\partial r}=\rho_a^{-1/2}\left(\frac{\partial u}{\partial r}+\kappa_au\right) \tag{39}$$

on $r=a$. Substituting Eqs. (37) and (38), making use of the orthogonality of the Legendre polynomials, gives

$$j_n(k_0a)+A_nh_n(k_0a)=\sigma B_nu_n(a),$$

$$k_0\{j_n'(k_0a)+A_nh_n'(k_0a)\}=\sigma^{-1}B_n\{u_n'(a)+\kappa_au_n(a)\}$$

for $n=0,1,2,...$, where $\sigma=(\rho_a/\rho_0)^{1/2}$. These two equations can be solved for A_n and B_n

$$A_n\Delta=(k_0\sigma)^{-1}j_n(k_0a)\{u_n'(a)+\kappa_au_n(a)\}-\sigma j_n'(k_0a)u_n(a)$$

and $B_n\Delta=i(k_0a)^{-2}$, where

$$\Delta=\sigma h_n'(k_0a)u_n(a)-(k_0\sigma)^{-1}h_n(k_0a)\{u_n'(a)+\kappa_au_n(a)\}.$$

B. A homogeneous sphere with an inhomogeneous coating

Suppose that the sphere $r<a$ consists of a homogeneous core $r<b$ (with density ρ_c and sound speed c_c) with an inhomogeneous concentric coating, $b<r<a$.

In the coating, we can write

$$u(r,\theta)=\rho_0^{1/2}c_0^2\sum_{n=0}^{\infty}(2n+1)\times i^n\{B_n u_n(r)+C_n v_n(r)\}P_n(\cos\theta),$$

where $u_n(r)$ and $v_n(r)$ are solutions of Eq. (36). We suppose that $u_n(r)$ is regular at $r=0$, whereas $v_n(r)$ is singular at $r=0$. In the homogeneous core, the pressure field is

$$p_c(r,\theta)=\rho_0 c_0^2\sum_{n=0}^{\infty}(2n+1)i^n D_n j_n(k_c r)P_n(\cos\theta),$$

where $k_c=\omega/c_c$.

We have to enforce two transmission conditions at $r=a$ and two at $r=b$. Let

$$\rho_b=\lim_{r\to b+}\rho(r) \text{ and } \kappa_b=\lim_{r\to b+}[\rho'(r)/\rho(r)].$$

Then, the interface conditions are Eq. (39) and

$$p_c=\rho_b^{1/2}u \text{ and } \rho_c^{-1}\frac{\partial p_c}{\partial r}=\rho_b^{-1/2}\left(\frac{\partial u}{\partial r}+\kappa_b u\right)$$

on $r=b$. These four conditions can be used to determine A_n, B_n, C_n, and D_n in a straightforward way.

V. EXPONENTIAL VARIATIONS IN ρ

Let us assume specific functional forms for $\rho(r)$ and $k(r)=\omega/[c(r)]$, namely

$$\rho(r)=\rho_1 e^{2\beta r} \text{ and } [k(r)]^2=k_1^2+2\alpha r^{-1}. \tag{40}$$

Here, ρ_1, β, k_1^2, and α are four adjustable constants. We find that Eq. (36) becomes

$$u_n''+2r^{-1}u_n'+[k_1^2-\beta^2+2(\alpha+\beta)r^{-1}-n(n+1)r^{-2}]u_n=0. \tag{41}$$

Equation (41) has a regular singularity at $r=0$, an irregular singularity at $r=\infty$, and no others. Therefore, it can be transformed into the confluent hypergeometric equation. Make the substitution

$$u_n(r)=r^{-1}w_n(x) \quad \text{with} \quad x=\delta r$$

in Eq. (41), giving

$$w_n''(x)+\left[\frac{k_1^2-\beta^2}{\delta^2}+\frac{2}{x}\frac{\alpha+\beta}{\delta}-\frac{n(n+1)}{x^2}\right]w_n(x)=0, \tag{42}$$

where δ is a parameter at our disposal. There are now three cases, depending on the relative sizes of k_1^2 and β^2.

A. Case I ($k_1^2>\beta^2$)

Choose $\delta^2=k_1^2-\beta^2$ and set $\eta=-(\alpha+\beta)/\delta$. Then, Eq. (42) becomes

$$w_n''(x)+[1-2\eta x^{-1}-n(n+1)x^{-2}]w_n(x)=0,$$

which is the *Coulomb wave equation* (Abramowitz and Stegun, 1965, Chap. 14). Its general solution is

$$w_n(x)=A_n F_n(\eta,x)+B_n G_n(\eta,x),$$

where A_n and B_n are arbitrary constants, F_n is the regular Coulomb wave function (bounded at $x=0$), and G_n is the irregular Coulomb wave function. These functions arise in nuclear physics (Biedenharn and Brussard, 1965, Chap. 3, Sec. 4).

Unsurprisingly, Coulomb wave functions are wavelike, in the sense that

$$G_n(\eta,x)+iF_n(\eta,x)\sim e^{i(x-\varphi)} \quad \text{as} \quad x\to\infty,$$

where $\varphi=\eta\log 2x+\frac{1}{2}n\pi-\sigma_n$ and σ_n is known [Abramowitz and Stegun, 1965, Eq. (14.6.5)]. Moreover,

$$F_n(0,x)=xj_n(x) \quad \text{and} \quad G_n(0,x)=-xy_n(x),$$

where j_n and y_n are spherical Bessel functions, so that the known solutions for homogeneous media are recovered.

B. Case II ($k_1^2=\beta^2$)

Choose $\delta=8(\alpha+\beta)$, and then Eq. (42) becomes

$$w_n''(x)+[\tfrac{1}{4}x^{-1}-n(n+1)x^{-2}]w_n(x)=0, \tag{43}$$

which is related to Bessel's equation; the general solution of Eq. (43) is

$$w_n(x)=\sqrt{x}\{A_n J_{2n+1}(\sqrt{x})+B_n Y_{2n+1}(\sqrt{x})\}.$$

C. Case III ($k_1^2<\beta^2$)

Choose $\delta^2=4(\beta^2-k_1^2)$ and set $\kappa=2(\alpha+\beta)/\delta$ and $\mu=n+\frac{1}{2}$. Then, Eq. (42) becomes

$$w_n''(x)+[-\tfrac{1}{4}+\kappa x^{-1}+(\tfrac{1}{4}-\mu^2)x^{-2}]w_n(x)=0, \tag{44}$$

which is known as *Whittaker's equation*. Its general solution is given by

$$w_n(x)=A_n M_{\kappa,\mu}(x)+B_n W_{\kappa,\mu}(x),$$

where $M_{\kappa,\mu}$ and $W_{\kappa,\mu}$ are Whittaker functions; these are discussed by Whittaker and Watson (1927, Chap. 16), by Erdélyi *et al.* (1953, Sec. 6.9), by Abramowitz and Stegun (1965, Chap. 13), and by Buchholz (1969). The occurrence of Whittaker functions is a little surprising, because these functions do not exhibit wavelike behavior. Thus

$$M_{\kappa,\mu}(x)\sim x^{-\kappa}e^{x/2} \quad \text{and} \quad W_{\kappa,\mu}(x)\sim x^{\kappa}e^{-x/2}$$

as $x\to\infty$. Moreover [Buchholz, 1969, Sec. 2, Eqs. (11a) and (29a)]

$$M_{0,\mu}(x)=\sqrt{\pi x}I_\mu(x) \text{ and } W_{0,\mu}(x)=\sqrt{x/\pi}K_\mu(x),$$

where I_μ and K_μ are modified Bessel functions. We remark that Whittaker functions also occur when solving the steady-state heat-conduction equation, $\operatorname{div}[k(r)\operatorname{grad}u]=0$, when $k(r)$ varies exponentially with r; see Martin (2002). The solutions described above can be inserted into the formulas obtained by the method of separation of variables in Sec. IV.

VI. GAUSSIAN SPHERES

One drawback of the functional forms in Eq. (40) is that the corresponding sound speed satisfies $c(0)=0$ (unless $\alpha=0$), so that Eq. (40) may not be suitable for an inhomogeneous sphere. (This objection does not apply if the sphere has a homogeneous core, as described in Sec. IV B.)

As an alternative, we can make progress by supposing that the density is a Gaussian and that $k^2(r)$ is linear in r^2. Thus, we suppose that

$$\rho(r)=\rho_1 e^{-\beta r^2} \text{ and } [k(r)]^2=k_1^2+\gamma r^2, \tag{45}$$

where ρ_1, β, k_1^2, and γ are adjustable constants. We find that Eq. (36) becomes

$$u_n''+2r^{-1}u_n'+[(\gamma-\beta^2)r^2+(k_1^2-3\beta)-n(n+1)r^{-2}]u_n=0.$$

To simplify this equation, make the substitution $u_n(r)=r^{-3/2}w_n(x)$ with $x=\delta r^2$, where δ is a disposable parameter. This gives

$$w_n''(x)+\left[\frac{\gamma-\beta^2}{\delta^2}+\frac{k_1^2-3\beta}{4\delta x}-\left\{n(n+1)-\frac{3}{16}\right\}x^{-2}\right]w_n(x)=0. \tag{46}$$

As in Sec. V, there are now three cases, depending on the sign of $\gamma-\beta^2$. For example, we obtain the Coulomb wave equation if $\gamma>\beta^2$ and Whittaker's equation if $\gamma<\beta^2$. Explicit solutions follow readily, but are not recorded here.

VII. CONCLUSIONS

In this paper, we have done two things. First, we have given a derivation of Bergmann's equation for sound waves in inhomogeneous media, where both the ambient density and sound speed can vary with position (but not time). Second, we have studied the scattering of waves by an inhomogeneous sphere; for certain exponential variations in the density, such scattering problems can be solved by the method of separation of variables, where the radial dependence involves some less-familiar but well-studied special functions. Perhaps the main value of these solutions is to provide nontrivial benchmarks against which numerical schemes (based, for example, on volume integral equations) can be tested.

ACKNOWLEDGMENTS

Thanks are due to C. J. Chapman and J. A. DeSanto for their constructive comments on the derivation in Sec. II A.

Abramowitz, M., and Stegun, I. A. (editors) (**1965**). *Handbook of Mathematical Functions* (Dover, New York).
Ahner, J. F. (**1977**). "Scattering by an inhomogeneous medium," J. Inst. Math. Appl. **19**, 425–439.
Batchelor, G. K. (**1967**). *An Introduction to Fluid Mechanics* (University Press, Cambridge).
Bergmann, P. G. (**1946**). "The wave equation in a medium with a variable index of refraction," J. Acoust. Soc. Am. **17**, 329–333.
Biedenharn, L. C., and Brussard, P. J. (**1965**). *Coulomb Excitation* (Clarendon, Oxford).
Brekhovskikh, L. M. (**1960**). *Waves in Layered Media* (Academic, New York).
Buchholz, H. (**1969**). *The Confluent Hypergeometric Function* (New York, Springer).
Colton, D. (**1978**). "The scattering of acoustic waves by a spherically stratified medium and an obstacle," SIAM (Soc. Ind. Appl. Math.) J. Math. Anal. **9**, 935–942.
Colton, D., and Kress, R. (**1978**). "The construction of solutions to acoustic scattering problems in a spherically stratified medium," Q. J. Mech. Appl. Math. **31**, 9–17.
Colton, D., and Kress, R. (**1979**). "The construction of solutions to acoustic scattering problems in a spherically stratified medium. II," Q. J. Mech. Appl. Math. **32**, 53–62.
Colton, D., and Kress, R. (**1992**). *Inverse Acoustic and Electromagnetic Scattering Theory* (Springer, Berlin).
DeSanto, J. A. (**1992**). *Scalar Wave Theory* (Springer, Berlin).
Erdélyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F. G. (**1953**). *Higher Transcendental Functions* (McGraw-Hill, New York), Vol. 1.
Frisk, G. V., and DeSanto, J. A. (**1970**). "Scattering by spherically symmetric inhomogeneities," J. Acoust. Soc. Am. **47**, 172–180.
Kai, L., and Massoli, P. (**1994**). "Scattering of electromagnetic-plane waves by radially inhomogeneous spheres: A finely stratified sphere model," Appl. Opt. **33**, 501–511.
Kleinman, R. E., and Martin, P. A. (**1988**). "On single integral equations for the transmission problem of acoustics," SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **48**, 307–325.
Kriegsmann, G. A., and Reiss, E. L. (**1983**). "Low frequency scattering by local inhomogeneities," SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **43**, 923–934.
Li, Y. L., Liu, C. H., and Franke, S. J. (**1990**). "Three-dimensional Green's function for wave propagation in a linearly inhomogeneous medium—the exact analytic solution," J. Acoust. Soc. Am. **87**, 2285–2291.
Lighthill, J. (**1978**). *Waves in Fluids* (University Press, Cambridge).
Martin, P. A. (**2002**). "On functionally graded balls and cones," J. Eng. Math. (in press).
Morse, P. M., and Ingard, K. U. (**1986**). *Theoretical Acoustics* (Princeton University Press, Princeton).
Newton, R. G. (**1982**). *Scattering Theory of Waves and Particles*, 2nd ed. (Springer, New York).
Ostashev, V. E. (**1997**). *Acoustics in Moving Inhomogeneous Media* (Spon, London).
Perelman, A. Y. (**1996**). "Scattering by particles with radially variable refractive indices," Appl. Opt. **35**, 5452–5460.
Pierce, A. D. (**1989**). *Acoustics* (Acoustical Society of America, New York).
Pierce, A. D. (**1990**). "Wave equation for sound in fluids with unsteady inhomogeneous flow," J. Acoust. Soc. Am. **87**, 2292–2299.
Sleeman, B. D. (**1980**). "Acoustic scattering by inhomogeneous media," Q. J. Mech. Appl. Math. **33**, 373–383.
Whittaker, E. T., and Watson, G. N. (**1927**). *A Course of Modern Analysis*, 4th ed. (University Press, Cambridge).
Wyatt, P. J. (**1962**). "Scattering of electromagnetic plane waves from inhomogeneous spherically symmetric objects," Phys. Rev. **127**, 1837–1843. Errata: **134** (1964) AB1.

An investigation of the influence of acoustic waves on the liquid flow through a porous material

Pietro Poesio, Gijs Ooms, and Sander Barake
J. M. Burgers Center, Delft University of Technology, Laboratory for Aero- and Hydrodynamics, Leeghwaterstraat 21, 2628 CB Delft, The Netherlands

Fred van der Bas
Shell International Exploration and Production, Technology Application and Research, P.O. Box 60, 2280 AB Rijswijk, The Netherlands

(Received 26 February 2001; revised 7 December 2001; accepted 31 January 2002)

An experimental and theoretical investigation has been made of the influence of high-frequency acoustic waves on the flow of a liquid through a porous material. The experiments have been performed on Berea sandstone cores. Two acoustic horns were used with frequencies of 20 and 40 kHz, and with maximum power output of 2 and 0.7 kW, respectively. Also, a temperature measurement of the flowing liquid inside the core was made. A high external pressure was applied in order to avoid cavitation. The acoustic waves were found to produce a significant effect on the pressure gradient at constant liquid flow rate through the core samples. During the application of acoustic waves the pressure gradient inside the core decreases. This effect turned out to be due to the decrease of the liquid viscosity caused by an increase in liquid temperature as a result of the acoustic energy dissipation inside the porous material. Also, a theoretical model has been developed to calculate the dissipation effect on the viscosity and on the pressure gradient. The model predictions are in reasonable agreement with the experimental data. [DOI: 10.1121/1.1466872]

PACS numbers: 43.20.Jr, 43.25.Nm [DEC]

I. INTRODUCTION

Fouling of the near-well bore region can decrease the oil recovery from an oil reservoir dramatically. Different techniques have been developed to remove this fouling; for instance, the injection of acids that dissolve the fouling particles. These techniques have undesirable side effects: they are costly, impose production shut-off, or have a negative impact on the environment. Therefore, in more recent years a new cleaning method has been proposed and investigated: treatment of the near-well bore region with high-frequency, high-energy acoustic waves. However, the results of this technique are rather variable. Acoustic cleaning causes an increase in the oil flow rate in only about 50% of the cases. In a number of cases it even has a negative effect on the flow rate. In order to optimize this technique it is essential to understand the physical mechanism of acoustic cleaning.

In this publication we go back even one step further. We first try to understand the influence of high-frequency acoustic waves on the flow of a liquid in a porous material without fouling. Only when we have understood this influence will we take the next step: to understand the effect of acoustic waves on fouling particles present in the liquid in the porous material. In the review article by Beresnev and Johnson,[1] several mechanisms have been proposed for the influence of high-frequency acoustic waves on the flow of a liquid in a porous material. They mention, for instance, the reduction in adherence between pore wall and liquid, acoustic streaming, acoustic cavitation, in-pore turbulence, viscosity decrease due to energy dissipation, etc. Aart and Ooms[2] have developed a model based on a peristaltic liquid transport caused by a deformation of the pore walls due to the traveling acoustic waves.

However, no certainty exists as to which mechanism is the correct one. We focus our investigation on Berea sandstone with permeability 100 to 300 mD and on acoustic frequencies below the critical frequency as defined by Biot.[3] We have carried out a detailed experimental and theoretical investigation of the influence of acoustic waves on the (laminar) liquid flow through a porous material. The results are given in this publication.

II. EXPERIMENTAL SETUP

A. Cores

The cores that are used for the experiments are cylindrically shaped Berea sandstone samples. They are very representative for the type of porous material in an oil reservoir. The length of the cores is 20 cm and the diameter 7.62 cm. The porosity is about 25%. The initial permeability is 100–300 mD for all samples.

B. Setup

A core is placed in a rubber sleeve to keep it fixed during the experiment. It is then placed in a steel vessel in which down-hole conditions are simulated (up to 150 bar and 100 °C). An acoustic horn is placed at one end of the core (see Figs. 1 and 2).

The high pressure in the vessel makes it possible to avoid cavitation (for pressures lower than 100 bar the influ-

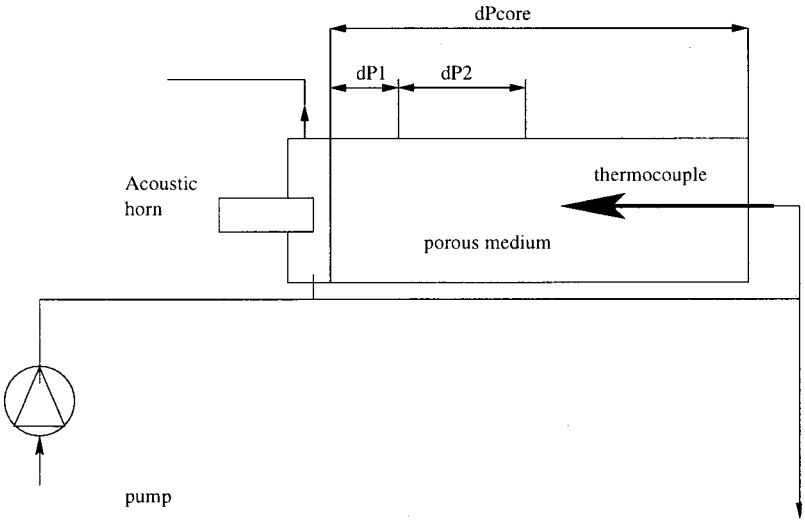


FIG. 1. Experimental setup with one temperature measurement.

ence of cavitation becomes noticeable). The space between the vessel and sleeve, which is filled with water, is then pressurized to 180 bar to make the rubber sleeve completely sealed off. A constant flow of brine (a 2% KCl solution in water) through the core is generated by means of a pump. So, because of mass conservation, the local superficial velocity in the core is constant. We use brine instead of plain water to avoid colloidally induced migration of fines. A flow in both directions is possible. This allows us to have the liquid flow and the acoustic wave propagating in the same direction or in opposite directions. There are four pressure measurements, two along the core (at 2.54 and 10.70 cm) and two at both ends of the core (see Figs. 1 and 2). dP_1 is the pressure drop over the first part of the core sample, dP_2 the pressure drop over the middle part, and dP_c the total pressure drop over the core. The pressure drop over the third part (dP_3) can be calculated in the following way: $dP_3 = dP_c - (dP_1 + dP_2)$. For the first experiments a thermocouple was placed in the middle part of the core, approximately 10 cm deep, from the side wall. A hole was drilled in the core and the thermocouple placed inside (see Fig. 1). Liquefied metal was poured in the hole to fix the thermocouple. The aim was to measure the liquid temperature in the middle part of the core during the experiments. During the first experiments we found a very interesting effect with respect to the temperature. In order to investigate this effect in more detail, we installed two thermocouples at the side wall of the porous medium, through the rubber sleeve (see Fig. 2). This new temperature measurement method is better than the first one, as the probes do not disturb the flow field. Results with both measurement methods will be reported in this publication. Calculations show that the temperature profile is uniform within the core. Also, the temperature in front of the core can be measured. The data are sent to a digital data recorder and processed on a computer.

C. Ultrasonic equipment

The ultrasonic equipment consists of

(i) a converter, which converts electricity into mechanical vibrations of a piezoelectric element;
(ii) an amplifier, which is used to set the amplitude of the vibrations; and

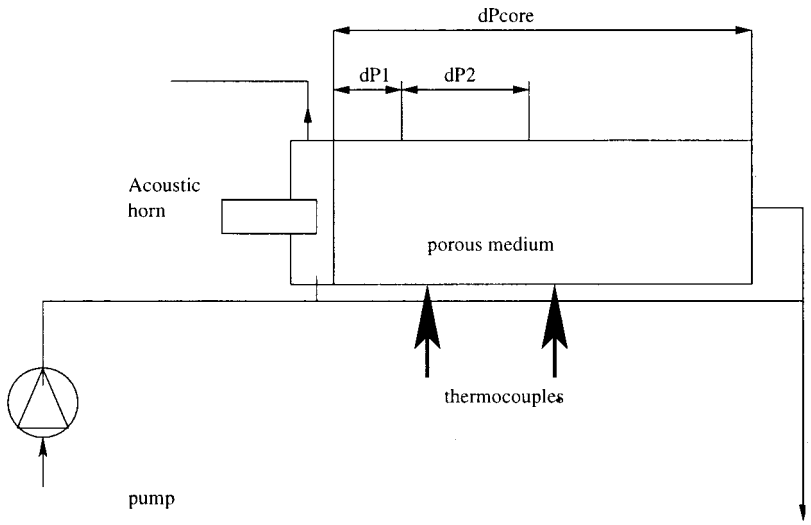


FIG. 2. Experimental setup with two temperature measurements.

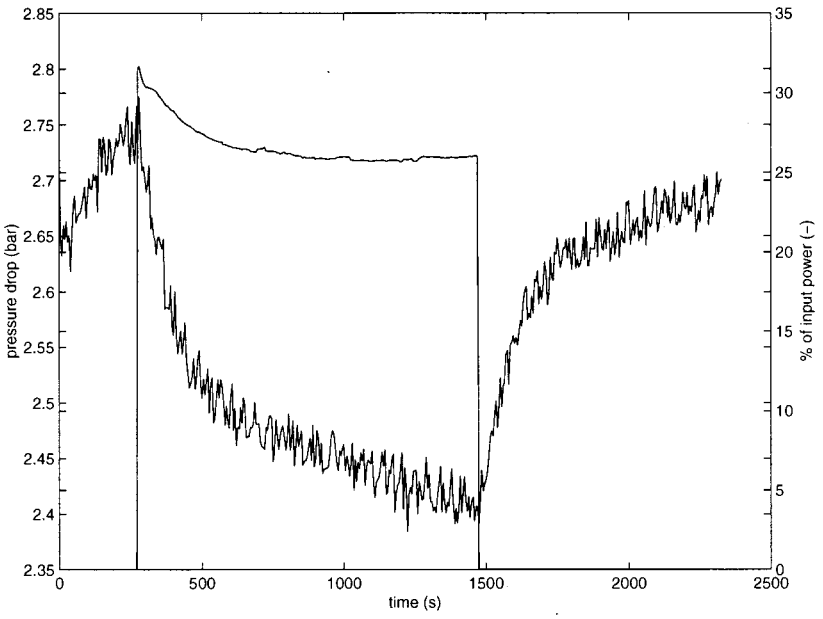


FIG. 3. Pressure drop over the middle part of the core (bold line) and percentage of power input (narrow line) as a function of time.

(iii) an ultrasonic horn, which concentrates the mechanical vibrations onto the front side of the core sample. Two acoustic horns are applied: a Branson module PGA 220 (a 20-kHz horn with maximum power output of 2000 W) and a Branson module PGA 470 (a 40-kHz horn with maximum power output of 700 W). The power output can be selected as a percentage of the maximum.

D. Microphone

A microphone is placed at the end side of the core. It is used to measure the amplitude of the acoustic signal after passage through the core. In this way the damping of the signal was determined during the experiments.

III. EXPERIMENTAL PROCEDURE AND RESULTS

A. Experimental procedure

A new core is used for each new series of experiments. Before performing experiments, the following steps are taken.

(i) The core is flushed with CO_2 to expel all the air in the porous material. The gas flow is kept on for several minutes.

(ii) After the gas flow is stopped a constant flow rate of the brine is started: the high pressure ensures that all the CO_2 dissolves into the brine.

(iii) The three pressure drops, the temperature in the middle part of the core (or at the sidewall of the core)

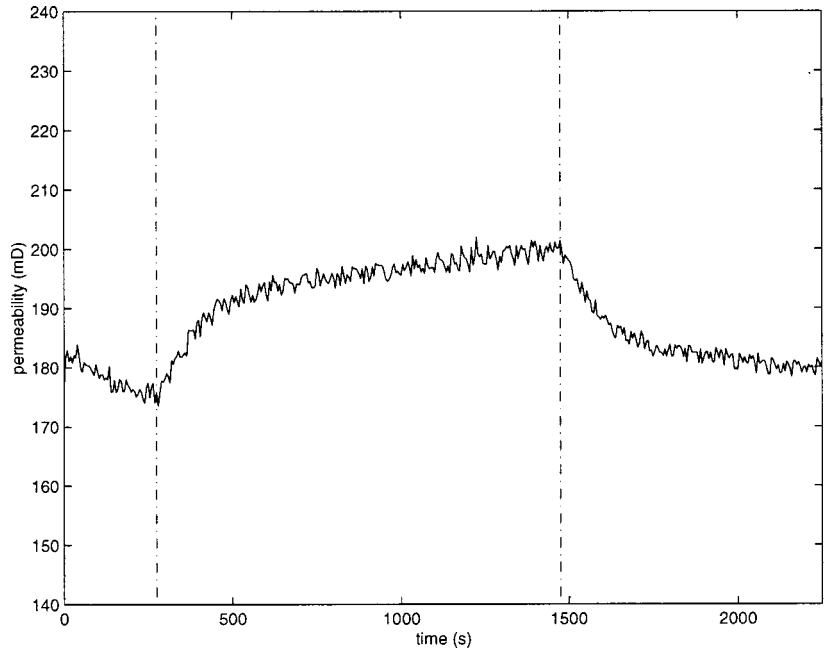


FIG. 4. Permeability as a function of time calculated assuming a constant temperature.

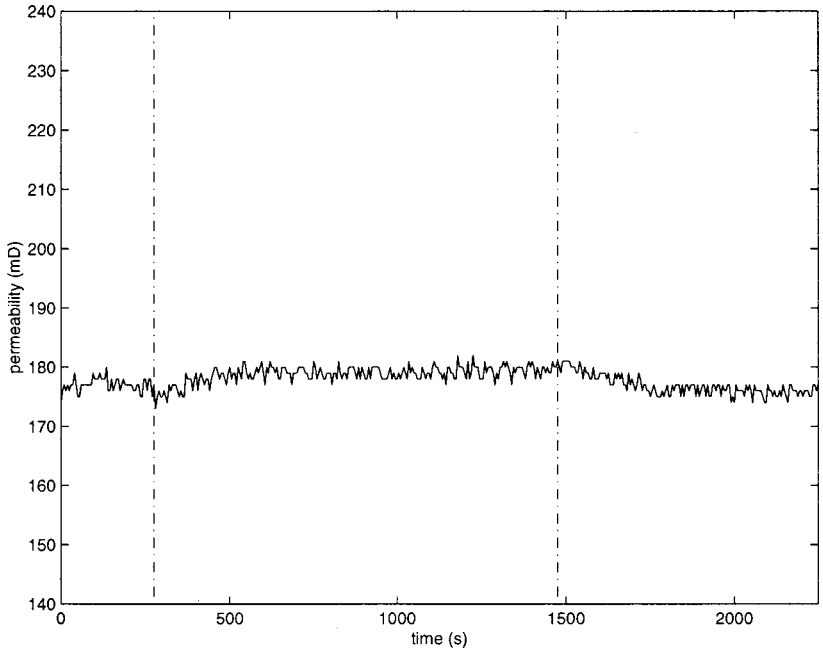


FIG. 5. Permeability as a function of time calculated using the measured temperature.

and in front of the core, and the constant flow rate of the brine are recorded.

(iv) The flow rate is increased up to the maximum value that the pump can deliver to detach and remove possible fines that are present in the core sample.

B. Experimental results

As an example of our experimental results, we show in Fig. 3 the pressure drop over the middle part of the core dP_2/dx as a function of time during one of the experiments: x is the distance to the front end of the core.

Also, the applied acoustic power is given as a function of time. In this experiment the acoustic frequency is 20 kHz, which is significantly lower than the critical value (230 kHz). The measurement of the pressure gradient over the middle part is accurate, as there is no influence of end effects of the core. For this reason we will leave out of consideration in this paper results for the pressure gradients dP_1/dx and dP_3/dx over the end parts. It can clearly be seen in Fig. 3 that during application of acoustic power the pressure gradient over the middle part of the core decreases. During this experiment the (constant) liquid flow and the acoustic wave propagation were in opposite directions. The flow rate was 75 ml/min.

From the measured flow velocity and pressure gradient as a function of time we have calculated, neglecting the influence of the temperature change on the liquid viscosity, the permeability of the porous material in the middle part of the

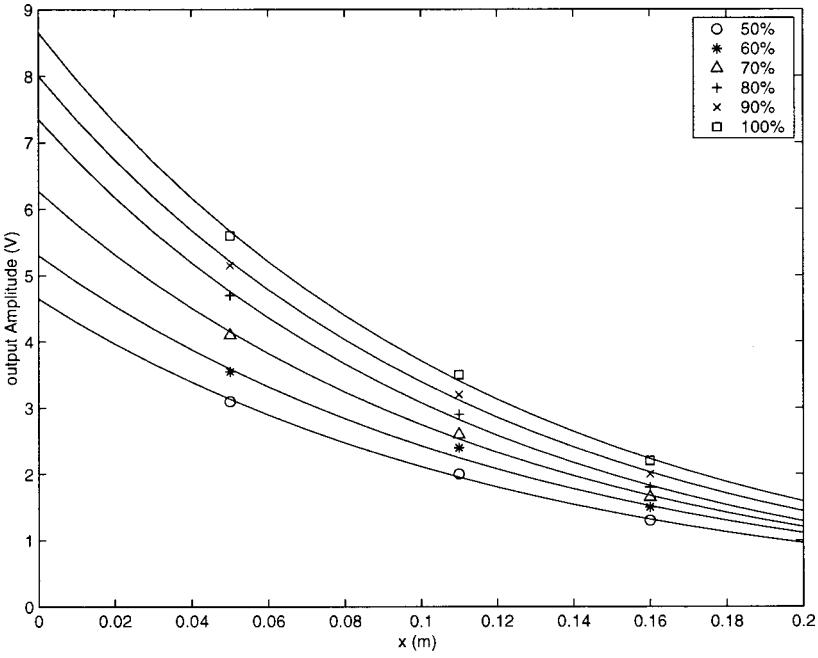


FIG. 6. Output signal as a function of core length. Symbols refer to the percentage of maximum power output. Conditions: $T=20\,^{\circ}\mathrm{C}$, $p_{\mathrm{pore}}=120$ bar, $p_{\mathrm{conf}}=185$ bar.

TABLE I. Parameters used for the calculation shown in Figs. 8 and 9.

Parameter	Value
α_{fast}	8.3 m^{-1}
α_{slow}	182 m^{-1}
Intensity of incident wave	0.7 kW
Permeability	200 mD
Porosity	25%

core sample as function of time by applying Darcy's law. This law is given by

$$\frac{dp}{dx} = -\frac{\mu}{K} v, \tag{1}$$

in which K is the permeability, μ the superficial liquid viscosity, and v the liquid velocity. As mentioned, the viscosity was considered constant and computed at the initial temperature (30 °C), measured in the core before the experiment. In Fig. 4 the calculated permeability for the middle part is shown as a function of time. It seems as if the high-frequency acoustic waves cause a significant increase of the (effective) permeability.

However, the liquid temperature has an influence on the liquid viscosity. Therefore, we repeated the calculation of the permeability of the middle part of the core, but this time taking the (known) influence of the temperature (measured in the middle part of the core) on the viscosity into account. The result is shown in Fig. 5, which represents the same experiment as in Fig. 4. The relationship between temperature and viscosity that we used can be found in the form of a table in Zaytsev and Aseyev.[4] As can be seen, there is no longer an influence of the high-frequency acoustic waves on the permeability. The permeability remains nearly constant. The pressure drop decrease observed during the application of acoustic waves is solely due to the decrease of the liquid viscosity, caused by the temperature increase of the liquid as a result of the acoustic energy dissipation in the porous material. This holds for all our experiments; both for the case where the acoustic wave propagation and liquid flow are in opposite directions and when they are in the same direction. In the following, we will compare this result with theoretical predictions.

IV. DAMPING COEFFICIENT

Before we can make theoretical predictions concerning the influence of acoustic wave propagation on the liquid flow through a porous material, we must know the damping of the wave inside the material. The incident acoustic wave is split in the porous medium into a slow wave and a fast wave. The damping of the slow wave is very fast and is known. However, the damping of the fast wave is much more difficult to predict. Biot's theory[3] predicts a damping coefficient for the fast wave which is too low for many porous materials. To the best of our knowledge, no models are available that can predict a damping coefficient for the fast wave which is in agreement with the experimental measurements. This is why we decided to determine this damping coefficient for our core samples experimentally. In order to measure the attenuation we used cores of different lengths (5, 11, and 16 cm) and measured the damping as a function of the core length with the aid of the microphone. Since we were not able to move the microphone (it is embedded in the plunger at the end of the core), the void space (in the case of the shorter cores) had to be filled in with a substitute material. We used Peek (a plastic). It can be assumed that the losses in the filling material are small compared to the losses in the porous material itself. We show an example of our experiments in Fig. 6. In this case, a brine-saturated core was used and the experiment was carried out at room temperature. For each core length, six different power inputs were used. As expected, each set of points could be interpolated by means of an exponential curve. The averaged damping coefficient was calculated to be 8.3 m^{-1}.

V. THEORETICAL CALCULATION

We will now calculate the influence of high-frequency acoustic waves on the flow in a porous medium, and in particular pay attention to the influence of the heat generation due to the dissipation of the acoustic waves on the flow. We first calculate the temperature distribution in the core as a function of time, taking into account the acoustic energy dissipation and the convective heat transport due to the liquid flow. We assume that the solid and the fluid are in thermal equilibrium: there is no heat transfer either from liquid to solid, or vice versa. To check this hypothesis we made use of the method proposed by Kaviany.[5] We may assume that the radial temperature distribution is uniform, because the rubber sleeve is a good insulating material. So, we consider the temperature to be a function of the axial coordinate only. The following heat balance equation is then satisfied:

$$[\rho_f c_{p,f}\phi + (1-\phi)\rho_s c_{p,s}]\frac{\partial T}{\partial t} + \rho_f c_{p,f} v \frac{\partial T}{\partial x} = \kappa \frac{\partial^2 T}{\partial x^2} + S_{\text{heat}}, \tag{2}$$

in which T is temperature, $c_{p,f}$ and $c_{p,s}$ are the specific heat for liquid and solid, respectively, ρ_f and ρ_s are the liquid and solid density, respectively, and v the superficial velocity of the liquid. ϕ is porosity. κ is the effective diffusion coefficient: for simplicity it is taken equal to the average of the values for the fluid and solid. In Eq. (2) we did not include a term representing the heat transfer from the core sample via the side wall and the rubber sleeve to the water in the surrounding vessel. However, the heat conductivity of the rubber sleeve is very small and the heat loss to the water is (according to our calculation) therefore negligible compared to the convective heat transfer. Only when the convective heat transfer is absent ($v=0$) will the heat transfer through the side wall become important. The source term in Eq. (2) may be written

$$S_{\text{heat}} = 2\alpha_{\text{fast}} I_{\text{fast},0} e^{-2\alpha_{\text{fast}} x} + 2\alpha_{\text{slow}} I_{\text{slow},0} e^{-2\alpha_{\text{slow}} x}. \tag{3}$$

$I_{\text{fast},0}$ and $I_{\text{slow},0}$ are the initial values (at the front end of the core) of the intensities of the fast wave and slow wave that are transmitted in the porous material as a result of the incident acoustic wave at the front end of the core. There is no transmitted shear wave, as the incident acoustic wave is perpendicular to the porous material surface at the front end.

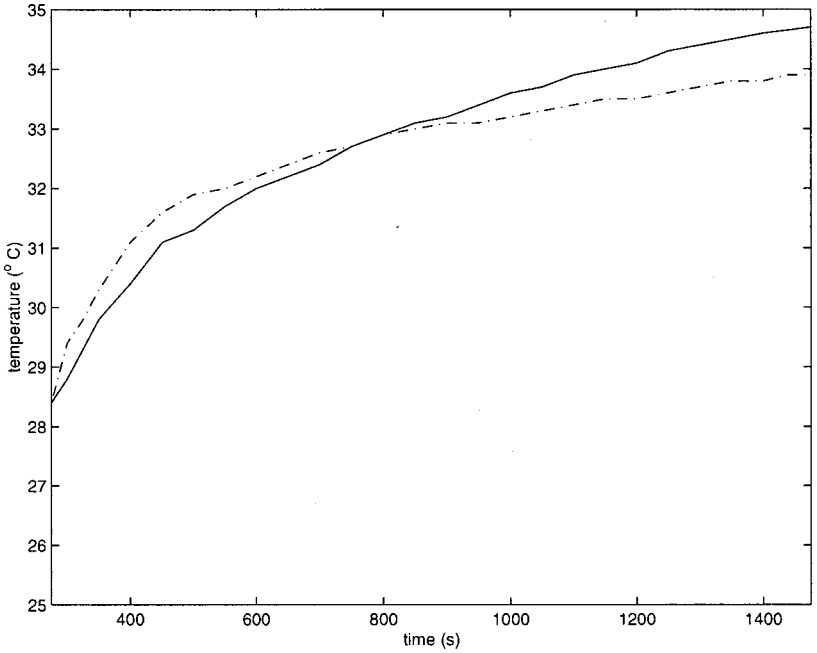


FIG. 7. Comparison between the measured temperature (continuous line) and the calculated value (dash-point line), using the setup in Fig. 1.

The damping coefficients for the fast wave and the slow wave are given by α_{fast} and α_{slow}.

To calculate the intensity of the fast and the slow waves from the intensity of the incident wave, we applied the procedure proposed by Wu *et al.*[6] using the Poynting vector for elastic waves in porous media. By definition the absolute value of the Poynting vector is the intensity of the wave.

As mentioned, it is well-known that the damping coefficient of the fast wave is considerably larger than the value calculated using Biot's theory.[3] Therefore, we used experimental values. For the experiments with the 2-kW horn, we used values available in the literature (Kelder[7]) because our microphone did not function properly for such a strong signal. Kelder measured experimentally the damping coefficient for the fast wave for Berea sandstone under different conditions. For the 700-W experiments, we measured the damping coefficient as described earlier. The values used are given in Table I.

With the parameter values from the table, we can solve the heat balance equation. In this way the temperature is calculated as function of time for a certain experiment. The predictions compare reasonably well with the measured temperature as a function of time. Two examples are given in Figs. 7 and 8. Now, assuming that the permeability is constant and taking into account the temperature dependence of the viscosity, we can calculate the pressure drop over the middle part of the core using Darcy's law. The result of this calculation for the experiment with the temperature change shown in Fig. 8 is given in Fig. 9. As can be seen from this figure, the experimental results and theoretical predictions

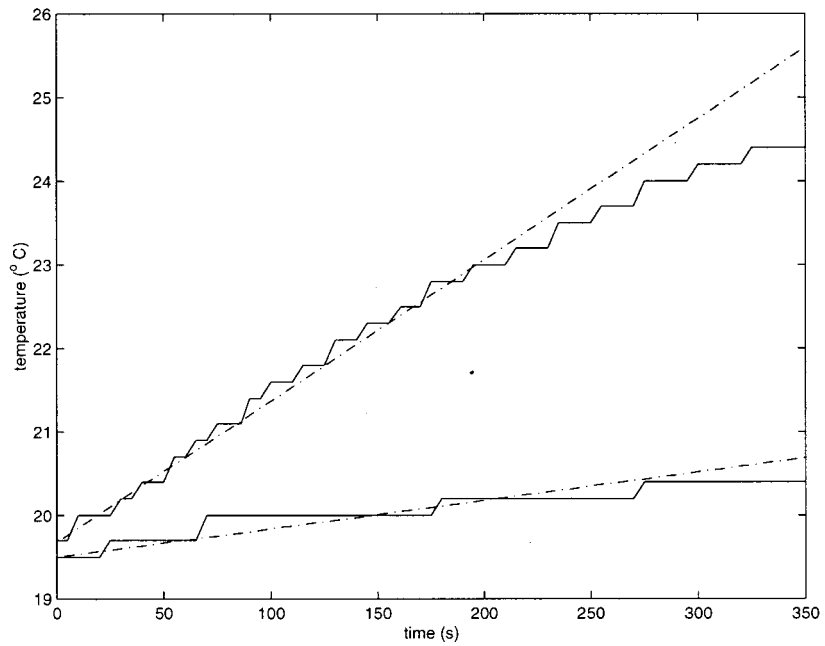


FIG. 8. Comparison between the measured temperature (continuous line) and the calculated value (dash-point line), using the setup in Fig. 2.

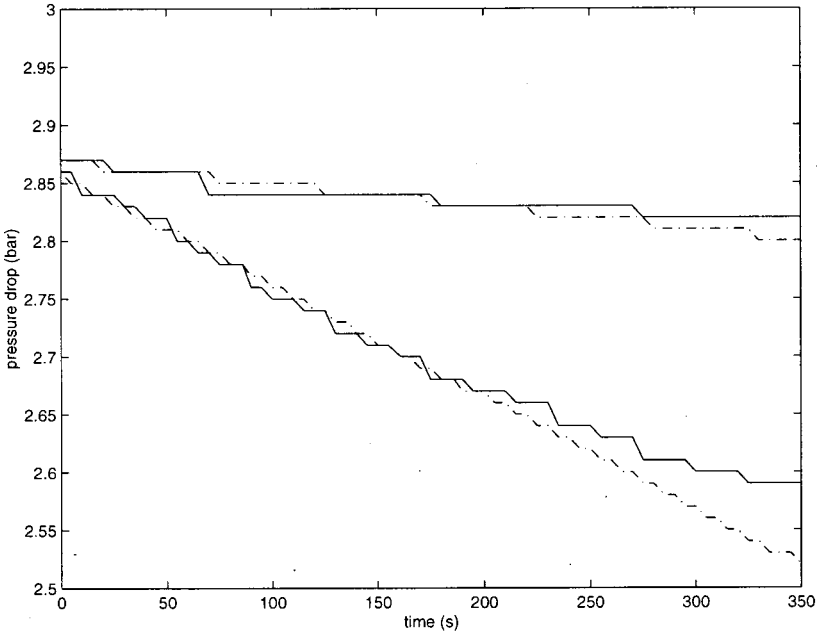


FIG. 9. Comparison between the measured pressure gradient (continuous line) and the calculated value (dash-point line), using the setup in Fig. 2.

are in reasonable agreement. This confirms again that the pressure drop decrease measured during the application of acoustic waves is due to the viscosity decrease caused by the heating of the liquid as a result of the acoustic energy dissipation.

VI. CONCLUSION

We investigated the influence of high-frequency acoustic waves on the flow in a porous material, in particular to find the main effect is responsible for the decrease on pressure drop when acoustic waves are applied. In the literature many different explanations are suggested. They are all possible, but there are no careful experimental validations. To that purpose we performed experiments on Berea sandstone and measured the pressure gradient and temperature in the middle part of the core samples at constant superficial liquid velocity. From our experimental data we conclude that the permeability does not change when acoustic waves are applied. The decrease in the pressure gradient can be explained completely by the decrease in fluid viscosity due to the dissipation of the acoustic energy. This fact is also confirmed by theoretical calculations. So, we have convincingly demonstrated that, beyond the effect of heating up the fluid and thereby changing its viscosity, there is nothing unusual.

ACKNOWLEDGMENTS

We wish to acknowledge Professor Dr. Ir. M. E. H. van Dongen and Dr. Ir. D. M. J. Smeulders for the fruitful discussions and suggestions throughout this work.

[1] I. A. Beresnev and P. A. Johnson, ''Elastic wave stimulation of oil production: A review of methods and result,'' Geophysics **59**, 1000–1017 (1994).
[2] A. C. T. Aarts and G. Ooms, ''Net flow of compressible viscous liquids induced by travelling waves in porous media,'' J. Eng. Math. **34**, 435–450 (1998).
[3] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-saturated porous medium. Low frequency range,'' J. Acoust. Soc. Am. **28**, 168–191 (1956).
[4] S. Zaytsev and V. Aseyev, *Properties of Aqueous Solution of Electrolytes* (CRC Press, Boca Raton, 1992).
[5] M. Kaviany, *Principles of Heat Transfer in Porous Media* (Springer, New York, 1995).
[6] K. Wu, Q. Xue, and L. Adler, ''Reflection and transmission of elastic waves from a fluid-saturated porous solid boundary,'' J. Acoust. Soc. Am. **87**, 2349–2358 (1990).
[7] O. Kelder, ''Frequency-dependent wave propagation in water-saturated porous media,'' Ph.D. thesis, TU Delft, 1998.

Real-time nondestructive evaluation of fiber composite laminates using low-frequency Lamb waves

Sergio H. Díaz Valdés and Costas Soutis[a)]
Department of Aeronautics, Imperial College of Science, Technology, and Medicine, Prince Consort Road, London SW7 2BY, United Kingdom

(Received 16 March 2000; revised 12 July 2001; accepted 29 January 2002)

Amid the nondestructive evaluation techniques available for the inspection of composite materials, only a few are suitable for implementation while the component is in service. The investigation examines the application of Lamb waves at low-frequency–thickness products for the detection of delaminations in thick composite laminates. Surface-mounted piezoelectric devices were excited with a tone burst to generate elastic waves in the structure. Experiments were carried out on composite beam specimens where wave propagation distances over 2 m were achieved and artificially induced delaminations as small as 1 cm^2 were successfully identified. The feasibility of employing piezoelectric devices for the development of smart structures, where a small and lightweight transducer system design is required, has been demonstrated. The resonance spectrum method, which is based on the study of spectra obtained by forced mechanical resonance of samples using sine-sweep excitation, has been proposed as a technique for measuring the A_o Lamb mode phase velocity. The finite-element method was also used to investigate qualitatively the dynamic response of laminates to wave propagation. Several locations and spatial distribution of the actuators were examined showing the advantages of using transducers arrays for the inspection of large structures. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1466870]

PACS numbers: 43.20.Ks, 43.35.Zc, 43.35.Yb, 43.60.Qv [RAR]

I. INTRODUCTION

The ability to evaluate the integrity of a structure without removing its individual components has become an important technology challenge. Several nondestructive evaluation (NDE) methods exist and are used in composite structures, where visual inspection, radiography, ultrasonics, shearography, and thermography are among the most commonly used. Despite their wide use and improvement in the last decades, the majority of NDE methods is not suitable for implementation into a smart structure. Applications that require a probe to obtain data and scanning of large areas are disregarded as the basis for the development of self-diagnostic systems, since they need direct intervention of humans to perform the inspection. In principle, an in-service health monitoring system would imitate a biological system, where attached or built-in sensors continuously interrogate the structural integrity throughout the component's life. Therefore, techniques that can operate from fixed locations in the structure while inspecting large areas are prime candidates for the development of a structural integrity monitoring system (SIMS). Furthermore, by fixing the transducer, many variables affecting the reliability and repeatability of measurements are removed allowing the precise assessment of minute changes in structural behavior, which permit the early detection of damage occurrence.

An attractive technique for the development of a SIMS is the use of Lamb waves. Their application has long been acknowledged as a potential solution for large-area nondestructive inspection, as they are able to travel relatively long distances, allowing the material between transmitter and receiver to be interrogated.[1] Hence, a line scan is achieved with each pulse rather than the comparatively slower point-scanning performance of conventional ultrasonic technique. Fundamentally, this method involves the analysis of the transmitted and/or reflected wave after interacting with the test part at boundaries or discontinuities. The presence of damage is identified when the response signal of subsequent tests deviates from the reference response of the undamaged configuration taken earlier in the structure's life.

Lamb waves can be excited and detected by a variety of methods, such as the use of interdigital transducers (IDTs),[2,3] fine point contact transducers,[4] air-coupled ultrasonic transducers,[5] laser-generation methods,[6] and the widely employed angled Perspex wedge.[7] However, among these methods only the use of IDTs appears suitable for implementation in smart-structures applications, where a small and lightweight, permanently attached transducer system design is required. Still, IDTs present some limitations for the Lamb wave inspection of thick sections that are commonly employed in practical structures. When used in ultrasonic applications, piezoelectric materials are normally operated at its thickness-mode (d_{33}) resonance frequency, which is determined by the thickness of the element and the longitudinal wave velocity in the material. The thickness of piezoelectric elements of practical use varies from few microns to few millimeters; thus, the dynamic range of piezoelectric transducers extends from low MHz (0.5) for thick elements to a few hundred MHz for very thin films. Since Lamb wave velocity depends on both the excitation frequency and thickness of the plate combined in a frequency–thickness (fh) product, then the lower operational frequency of a transducer

a)Electronic mail: c.soutis@ic.ac.uk

imposes an upper limit on the thickness of the plate in which a Lamb wave can be excited below the cutoff fh product of high-order Lamb modes. This condition is important in the use of Lamb waves for NDE applications, since the excitation of a single Lamb mode favors signal interpretation. One example is frequencies less than 250 and 100 kHz, respectively, would be necessary in order to have only the fundamental Lamb modes (A_o and S_o) propagating in typical composite aerospace laminates, which have thickness ranging from 6 to 15 mm.[8] Likewise, if a 12-mm-thick steel plate is to be tested, this implies working at frequencies below 136 kHz.[7]

This work investigates the generation of Lamb waves for the NDE of composite laminates using surface-mounted piezoelectric elements of narrow strip shape, operated in the longitudinal mode (d_{31}). When voltage is applied to a bonded piezoelectric patch, it expands and contracts parallel to the surface, inducing a bending moment in the structure. If the voltage applied is a sinusoid of few cycles, it then generates a transient flexural wave whose transmission, propagation, and subsequent reflections at the specimen's boundaries can be analyzed and used to identify the size and location of damage. The advantage of using the longitudinal or radial modes (related to width and length, or diameter) of the piezoelectric element rather than the thickness mode is that the former can be excited at much lower frequencies, which allows the inspection of thicker laminates while keeping the fh product low, thus generating only fundamental Lamb modes.

The study first examines the dynamic response of narrow-beam specimens to sine-sweep excitation and employs the resonance spectrum method to perform wave velocity measurements.[9,10] The dispersion curves obtained confirm the effective generation of Lamb waves and provide the information needed to locate and characterize damage on composite beam specimens. Then, the finite-element method takes the investigation one step further, from beams to plates, and is used to predict wave generation and propagation in wide laminates.

II. EXPERIMENTAL PROCEDURE AND MEASUREMENTS

A. Materials and instrumentation

Tests were performed on narrow-beam specimens made of aluminum and composite material. The beams were supported on a cushioning platform (Styrofoam) to simulate free–free boundary conditions. The composite beam was obtained from a 24-ply (layer sequence, $[\pm 45°/0°/90°]_{3s}$) carbon fiber/epoxy laminate of size 660×570 mm. The laminate was fabricated using T300-924C prepreg tapes (Hexcel Composites). Individual test specimens 629×25 mm and 2.7 mm thick were cut from the laminate using a diamond-wheel saw. The elastic properties of the unidirectional ply were $E_{11}=162$ GPa, $E_{22}=11$ GPa, $\nu_{12}=0.34$, $G_{12}=5.7$ GPa, and density $\rho=1536$ Kg/m^3.[11] The dimensions of the aluminum specimen were 814×16 mm and 3.3 mm thick. The beams were instrumented with two piezoelectric patches, 20×5 mm, made of commercial brass-backed piezoceramic resonators (Maplin Electronics Ltd.), used as actuator and sensor, respectively, and bonded near the beam end, as shown in Fig. 1. A thin layer of cyanoacrylate (Bostik superglue) was uniformly applied to the brass back and pressed firmly against the beam surface; the excess glue was removed with an absorbent cloth or tissue. Minimal surface preparation (dry and clean) was needed to ensure good adhesion. National Instruments' LABVIEW® signal-processing software and an analog-to-digital card (PCI-MIO-16E-1) were used in conjunction with a personal computer (Dell OptiPLex Gxa, Pentium II 300) to implement the data transmission/acquisition under an automated framework and to perform the sensor response analyses. The actuator was excited with a 10-V software-generated signal. Both sensor and actuator were connected directly to the data acquisition card.

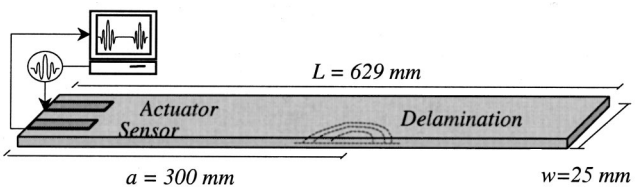


FIG. 1. Experimental setup of the composite beam specimen.

B. Phase velocity measurement

The wave velocity is the fundamental characteristic of a Lamb wave, since wave propagation may be analyzed by the variation of its velocity as a function of the fh product for each Lamb mode, as shown in Fig. 2. These relations (dispersion curves) are found by numerical solution of the Rayleigh–Lamb relation for wave propagation in isotropic plates, described comprehensively by Viktorov.[12] They can also be determined experimentally using the amplitude spectrum method[13] and the phase spectrum method.[14]

An alternative procedure to perform phase velocity measurements in the low-fh product range is the resonance spectrum method, which is based on the mechanical resonance response of beam specimens. In a narrow beam of length L, the particle velocity at resonance consists of a series of standing waves whose frequency correspond to the condition that an integral number of half-wavelengths fits in the sample. Thus, resonance exists when $2L=\lambda n$, where n is the harmonic integer and λ is the wavelength. Since the phase velocity is given by

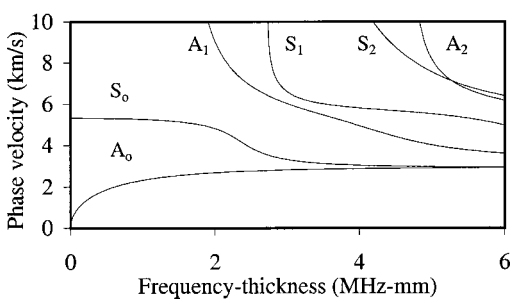


FIG. 2. Lamb wave phase velocity dispersion curves for aluminum.

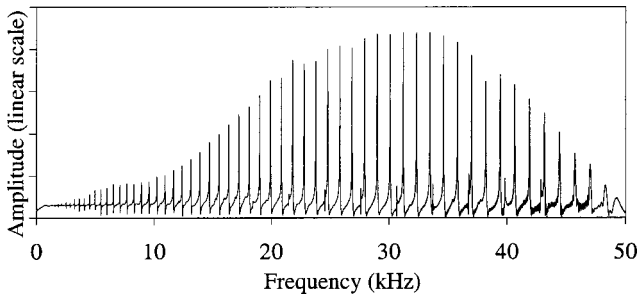


FIG. 3. Resonance spectrum of the response of the aluminum beam to sine-sweep excitation.

$$c_p = \lambda f, \tag{1}$$

then the nth mechanical resonance frequency can be expressed as

$$f_n = \frac{nc_p}{2L} \quad \text{or} \quad \omega_n = \frac{2\pi n c_p}{2L}. \tag{2}$$

Equation (2) gives, at discrete values of frequency (f_n), the information necessary to determine the variation of phase velocity as a function of frequency. In this fashion, the dispersion curve for the A_o Lamb mode in aluminum was obtained from the resonance spectrum of the aluminum beam. The actuator was excited with a 10-V sine sweep signal,[15] varying from 0.1 to 50 kHz in 0.2 s. The structural response captured with the sensor was sampled at a rate of 0.8 MHz and was then Fourier transformed to obtain the resonance frequencies shown in Fig. 3. The well-defined peaks shown in this figure correspond to the flexural modes of the beam, whereas the smaller peaks appearing above 20 kHz correspond to axial modes. Below 1 kHz, the resonance maxima were only poorly excited so that the modal number of each peak could not be indexed reliably. Therefore, the appropriate value of n was estimated by comparison with theoretical values of the natural frequencies of a uniform beam in transverse flexural vibration, using the expression[16]

$$f_i = \frac{\lambda_i^2}{2\pi L^2}\sqrt{\frac{EI}{m}}; \quad i = 1,2,..., \tag{3}$$

where m is the mass per unit length of the beam, E is the modulus of elasticity, I is the second moment of area, and λ_i is the solution of the characteristic equation for the imposed boundary conditions (free–free).

Figure 4 shows the dispersion curve for the A_o mode in aluminum obtained using the resonance frequency values from Fig. 3, substituted in Eq. (2). Also, the beam theory results derived using Eqs. (2) and (3) are presented in Fig. 4 along with the Lamb theory curve for aluminum, showing excellent correlation with the experimental results.

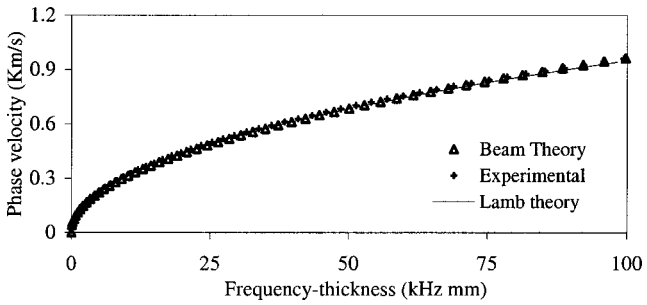


FIG. 4. Phase velocity in a 3.3-mm-thick aluminum beam measured using the mechanical resonance method. Beam and Lamb wave theory estimates are also presented.

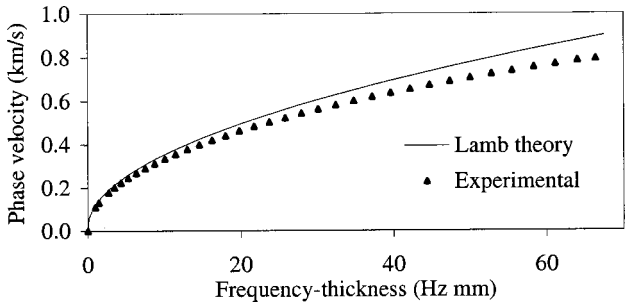


FIG. 5. Phase velocity in a 2.7-mm-thick $[\pm 45°/0°/90°]_{3s}$ composite beam measured using the mechanical resonance method. Lamb theory estimates are also presented.

The phase velocity of the A_o Lamb mode in the composite specimen was also measured using the resonance spectrum method. The sine signal used to excite the actuator was swept from 0.1 to 25 kHz in 0.2 s, and the structural response was sampled at a rate of 0.5 MHz. Figure 5 shows the experimental curve for the laminated beam along with the Lamb theory curve calculated using the average elastic properties of the orthotropic plate.[17]

The experimental data shown in Fig. 5 have the characteristic of the dispersion curve of the A_o Lamb mode, although the agreement between Lamb theory and experimental results is not quite as good as for the isotropic case, especially as the fh product increases (decreasing wavelength). The difference can be attributed to the fact that transverse shear deformation effects, which are neglected in the formulation of the classical plate theory, are significant in the case of laminated plates due to the relatively low transverse shear modulus.[18]

It can be seen from Fig. 4 that Lamb wave and beam theory can be used to successfully predict the dispersion curves of an isotropic material. However, for composite laminates there are no obvious analytical solutions, and numerical techniques are required to model such systems. In this study, we have proposed the resonance spectrum method as a reliable procedure for measuring low-frequency, long-wavelength flexural wave phase velocity. In addition, the results demonstrate that surface-mounted piezoelectric elements can effectively be used to generate Lamb waves in composite laminates. In the following section, this wave generation methodology will be employed for the NDE of composite laminates.

C. Delamination detection

Due to the relatively low interlaminar strength of composite laminates, damage in the form of delaminations can be easily introduced from low-velocity impacts during service, with the subsequent degradation of the mechanical properties of the laminate that can lead to the premature failure of structural components. In practice, composite laminates are designed to tolerate a certain degree of damage and it is often

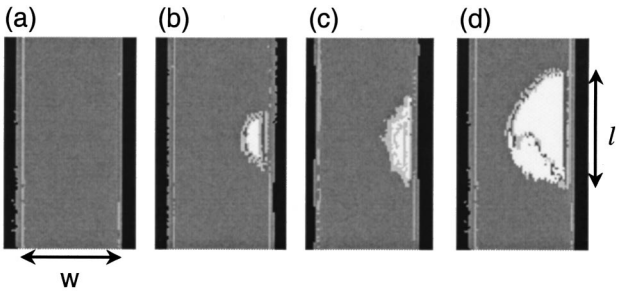


FIG. 6. Ultrasonic c-scans of the composite beam: (a) no delamination; (b) $A_d = 22$ mm^2; (c) $A_d = 47$ mm^2; (d) $A_d = 220$ mm^2.

only necessary to find relatively large defects such as 10–20-mm-diameter delaminations.[19] For instance, the typical size of a critical defect in a composite structure for a Harrier aircraft is approximately 20–25 mm.[8]

In order to study the ability of Lamb waves for delamination detection at low fh products, the composite beam was examined at different damage scenarios. The beam specimen was excited with short sinusoidal pulses rather than with continuous wave excitation, as was previously done for phase velocity measurements. The undamaged beam was tested and its response history was captured and kept as baseline. Then, a sharp and thin scalpel blade was inserted into the beam's midplane, 300 mm from the left end, as shown in Fig. 1. The blade was forced into the material, initially producing a small delamination whose dimension was increased each time the blade was forced into the midplane. In this manner the delamination area was gradually extended from a small incision located at one edge of the beam until it reached almost a full-width delamination. This type of artificially induced delamination is thought to better represent damage patterns observed in fatigue loading than beams with a full-width delamination of fixed length, as is commonly found in the literature. The damage area, A_d, was measured at every stage by conventional ultrasonic c-scan, as shown in Fig. 6.

Figure 7(a) shows the response of the undamaged composite beam when the actuator was excited with a 15-kHz sinusoidal pulse of 5.5 cycles modulated by a Hanning window. The response history shows the input pulse followed by another large wavelet, which is the first reflection from the opposite end of the beam. The time difference between the transmitted and reflected signals corresponds to the propagation distance of the wave, which is twice the length of the beam. The second and third reflection were also identified using longer acquisition times, suggesting a propagation range well over 2 m. However, the response signals shown in Fig. 7 were limited to capture the first reflection. It can also be observed that the shape of the wave changes as it propagates along the beam due to the dispersive nature of the A_o mode at this fh product.

The same test was performed after damage had been induced, thus monitoring the specimen response at different stages of delamination growth. In comparison to Fig. 7(a), Figs. 7(b)–(d) show an extra reflection between the input pulse and the first reflection from the end of the laminate. It can also be observed that the amplitude of the first reflection

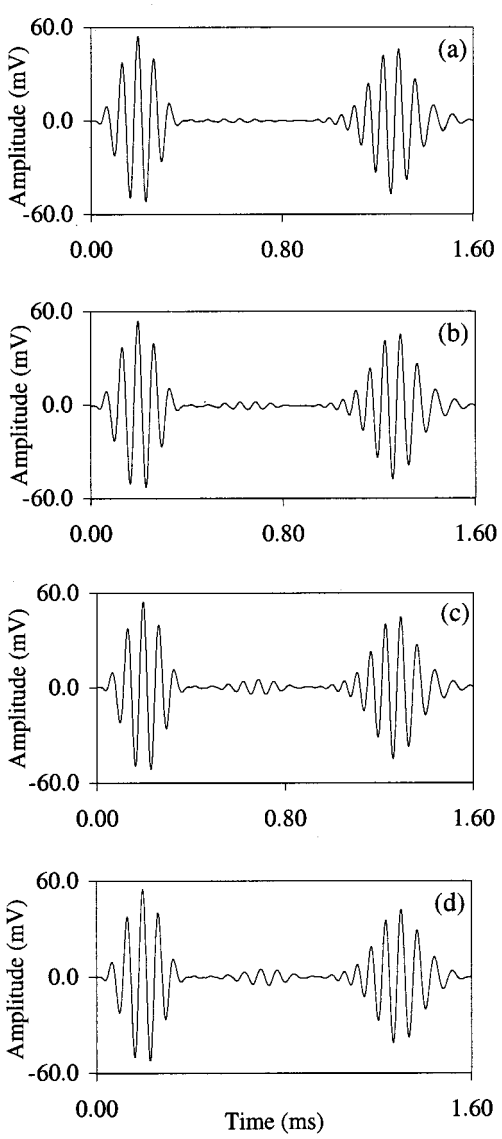


FIG. 7. Measured response of the composite beam at different stages of damage: (a) no defects; (b) $A_d = 22$ mm^2; (c) $A_d = 47$ mm^2; (d) $A_d = 220$ mm^2.

is affected (reduced) due to the presence of damage. These effects can be better appreciated in Fig. 8, which represents the arithmetic difference between Fig. 7(a) and each one of Figs. 7(a)–(d), respectively.

In Fig. 8 the reflection associated with the delamination (indicated by an arrow) and that of the end of the beam can be clearly identified. The increasing amplitude of the latter can be explained in part due to the differences in amplitudes between Fig. 7(a) and Figs. 7(b)–(d). However, it is mainly due to a shift in the time it takes for the input pulse to complete a round trip along the length of the specimen in the presence of damage. In general, the wave velocity changes in the delamination area since the wave has to travel through two regions (above and below the delamination) of smaller thickness than the undamaged laminate. In consequence, the fh product decreases and the phase velocity of the wave reduces according to the dispersion curves of the material (Fig. 5). Therefore, the time of travel of an A_o Lamb wave propagating in a specimen with delaminations is greater than

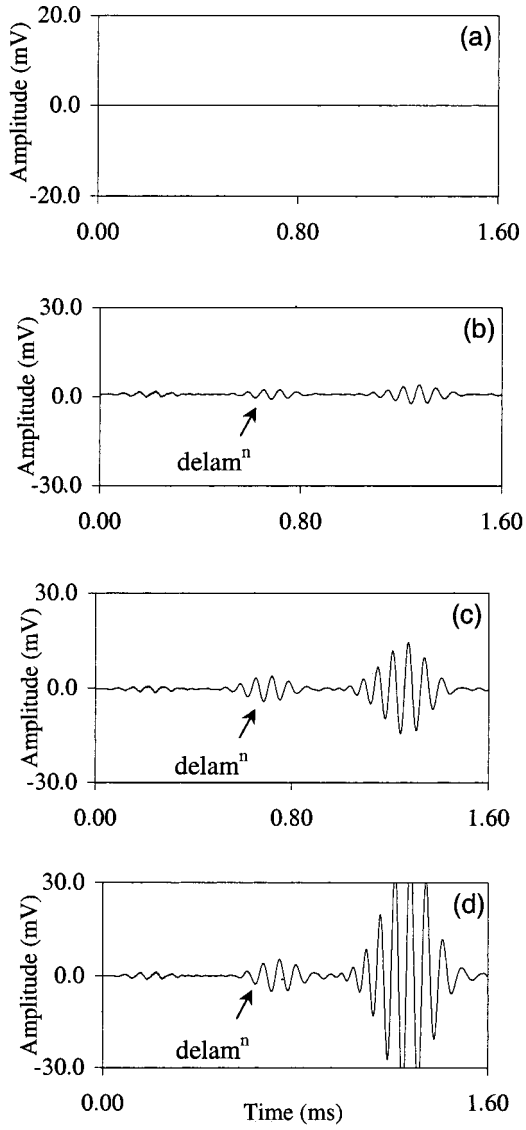


FIG. 8. The arithmetic difference between the response of the undamaged specimen and the response at different stages of damage: (a) no defects; (b) $A_d = 22\ \text{mm}^2$; (c) $A_d = 47\ \text{mm}^2$; (d) $A_d = 220\ \text{mm}^2$.

that of a wave traveling in an undamaged specimen.

The utility of subtracting the response histories, as in Fig. 8, is that the large-amplitude sections of the signals are eliminated, thus revealing small changes in the propagating wave such as variations in the amplitude and phase shifts. This procedure can be easily implemented in a health monitoring system, where damage occurrence would be identified by the appearance of ripples or wavelets in an otherwise straight-line signal [Fig. 8(a)] representative of the undamaged structure.

The sensitivity of the propagating wave to even the smallest delamination can be explained in terms of the relative dimension between wavelength and defect size. The wavelength of the excitation signal can be calculated from Eq. (1). The pulse center frequency used was 15 kHz, with plate thickness of 2.7 mm, which gives a frequency–thickness product of 40.5 kHz mm. At this fh value the phase velocity of the A_o mode obtained from Fig. 5 is close to 640 m/s, from which the wavelength is found to be $\lambda \approx 42$ mm. From the c-scans shown in Fig. 6, the l dimension (parallel to the length of the beam specimen) of the smallest delamination tested was 10 mm, showing that even though the delamination size is only $\approx 25\%$ of the wavelength, the system is capable of detecting it.

D. Delamination location

The position on the time scale of the reflection generated at the damage site and the reflection from the end of the laminate, shown in Fig. 7, can be used to estimate the location of the defect along the beam span. The time difference (Δt_L) between the maximum peak of the input pulse, and the maximum peak of the reflected signal from the end of the beam is about 1060 μs, which corresponds to twice the length of the specimen ($2L = 1258$ mm). Similarly, the time difference (Δt_{dam}) between the maximum peak of the input pulse and the maximum peak of the reflected signal from the damage site is about 490 μs, which corresponds to a round trip of the wave between the receiver and the delamination ($2a$). Substituting these values in Eq. (4)

$$a = \frac{L \Delta t_{\text{dam}}}{\Delta t_L}, \tag{4}$$

the location of damage (a) is found to be at 290 mm, which is a fair estimate of the actual location ($a = 300$ mm) of the artificially induced delamination shown in Fig. 1. This measurement is only approximate, since the shape of the wave packet does not remain the same during its propagation along the beam due to the dispersive nature of the A_o Lamb mode at this fh product.

These experiments demonstrate the potential use of Lamb waves at very low fh products for NDE applications, which along with simple and effective signal processing can be used for the development of an in-service, health monitoring technique capable of detecting delaminations in composite components. However, the specimens used were essentially narrow-beam elements where lateral wave-spreading effects are minimized. In the following section the wave propagation on wide plates is studied using the FE77 finite-element software.[20] Different actuation configurations were examined including the use of a linear array of actuators for the damage evaluation of large surfaces.

III. FINITE-ELEMENT ANALYSIS

A. Wave propagation in plates

The finite-element analysis was carried out in parallel with the experimental approach to study Lamb wave generation and propagation in quasi-isotropic beams and plates. Preliminary models were used to investigate qualitatively the interaction of the elastic wave at boundaries in order to predict the response of a surface-bonded sensor. A mesh in the $x-y$ plane was implemented to model the laminate using quadrilateral shell elements, where the nodes are defined on the mid-thickness of the shell and each node has both translational and rotational degrees of freedom. The x coordinate was aligned along the length direction and the z coordinate normal to the surface of the plate. The plate thickness is given as a geometrical parameter when the mesh is gener-

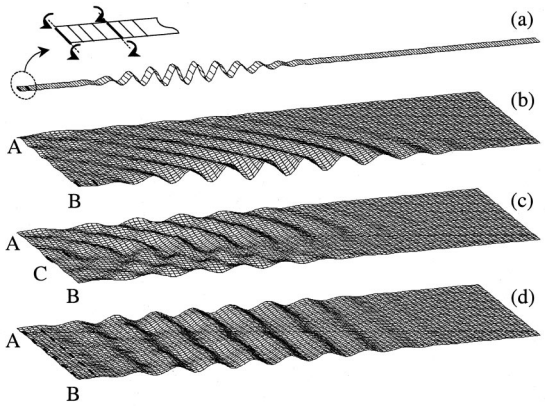


FIG. 9. The predicted response of 2-mm-thick laminates to a sinusoidal load applied at the locations indicated by bold lines on each mesh: (a) left end of the beam; (b) point B; (c) point C; (d) linear distribution along the AB boundary.

ated. The excitation of the plate using piezoelectric patches was modeled with uniformly distributed moments of opposite sign applied along two short, parallel nodal lines (actuator width), separated by a distance equivalent to the length of the piezoelectric actuator, as indicated with bold lines in Fig. 9. The excitation signal employed in all of the simulations was a 5.5-cycle, 20-kHz sinusoidal wave modulated by a Hanning window. It was defined in the code as a time history variation of the amplitude of the applied load. Then an explicit central difference scheme was employed in the MARCH module of the FE77, which carried out a step-by-step time-marching integration to solve the wave propagation simulation.

Figure 9 illustrates the predicted response of the specimens at a certain time step after being excited by the described load case. Figure 9(a) shows the propagation of the A_o Lamb mode along a 1000 mm$\times$30 mm beam. Likewise, Figs. 9(b) and (c) show the propagation of the A_o mode across the surface of a 500-$\times$200-mm plate excited with a single actuator located at its bottom left corner (point B) and at the center left edge (point C), respectively. In contrast to the wave propagation in the narrow beam, the plates present radial wave propagation with respect to the actuator location. This lateral wave spreading generates reflections at all of the plate boundaries, making it difficult to interpret the signal of the surface mounted sensor. Figure 9(d) shows the response of the same plate excited with a linear array of actuators distributed along its left edge (AB boundary). This arrangement produces a fairly uniform wavefront across the width of the plate, which is reflected by the right edge with minor lateral wave spreading. Such response favors the interpretation of the signal produced by a bonded sensor, since the input pulse and subsequent reflections are similar to those observed in the narrow beam and can be easily identified in the signal time history.

The advantages of using a linear array of actuators to produce a uniform wavefront can also be appreciated in Fig. 10, where the in-plane displacement history of a single node located at the center of the left edge is shown. The displacement histories shown correspond, respectively, to the plate response to each of the actuator distributions described

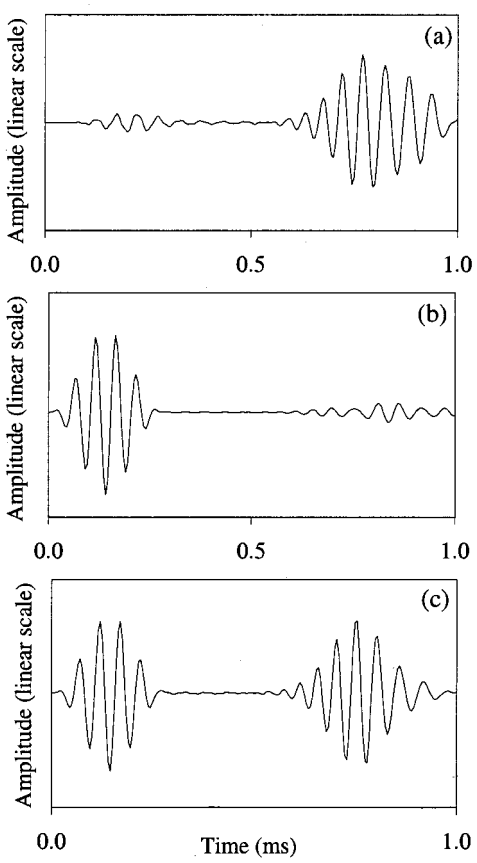


FIG. 10. The predicted displacement history of a node located at the center of the AB boundary. The plate was excited with a sinusoidal load applied at different locations: (a) bottom left corner; (b) center left edge; and (c) linear distribution along the AB edge.

above: actuator at the bottom left corner, point B; actuator at the center left edge, point C; and linearly distributed actuators along the left edge, AB boundary. It can be seen from Figs. 10(a) and (b) that the single-actuator scheme produces signal histories that do not favor signal interpretation, since the input pulse and first reflection are not clearly defined. Instead, Fig. 10(c) presents two well-defined sinusoids and a fairly flat signal between them, similar to that observed for the narrow-beam element, Fig. 7(a).

B. Damage detection

The advantages of using a linear array of actuators for the inspection of large areas are better exploited if the transducers are used in a pitch–catch mode (send–receive), thus employing the dual capabilities of piezoelectric materials as receiver and transmitter. This approach has been examined using a uniformly distributed sinusoidal load applied along the left edge of the plate model, the AB boundary in Fig. 11, while monitoring the in-plane nodal displacement along that same edge. Both undamaged and damaged configurations were simulated and their response compared to identify the effects of damage on the model behavior. Damage was simulated by degrading the material properties of a selected number of elements within the mesh. This reduction of the mechanical properties varied from zero stiffness to 50% of the elastic properties of the undamaged material, thus represent-

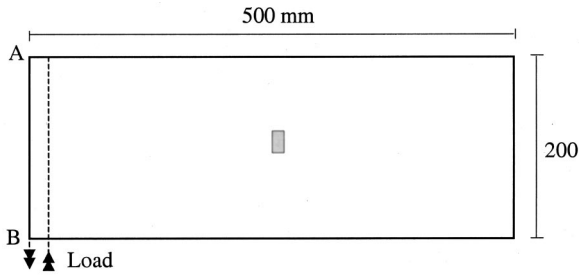


FIG. 11. Sketch of the finite-element model used in the calculations showing a central area with reduced material properties.

ing the cases of an open hole and damage due to low-velocity impact, respectively. Also, two different sizes of damage were studied, a square area of 10×10 mm and a rectangle of 20×10 mm, located at the center of the plate, as illustrated in Fig. 11. The dimensions of the damage sites are comparable to the wavelength ($\lambda\approx30$ mm) of the propagating pulse.

Figure 12 shows the response of the undamaged plate presented in a 3D plot of amplitude of in-plane displacement, time, and plate width. Similar to the experimental results, the first part of the response shows the input pulse followed by another large wavelet, which is the first reflection from the opposite end of the plate.

Figure 13(a) presents the response history of the plate containing damage in the form of a square open-hole. This displacement history is at first glance similar to that of the undamaged model, however, evident differences between the two time histories are revealed in Fig. 13(b), which represents the arithmetic difference between Figs. 12 and 13(a).

Likewise, Fig. 14(a) shows the response of the plate with the rectangular cutout, where the effects of damage are clearly shown; Fig. 14(b) presents the difference between this damage configuration and the undamaged plate. Qualitatively similar results to those shown in Figs. 13 and 14 were obtained in the case where the elastic properties in the damaged area were reduced by 50%, although the amplitude of the wave reflection generated at the damage site is lower than in the open-hole case.

The presence of structural discontinuities can be easily inferred from Figs. 13(b) and 14(b). In addition, the location of damage can be estimated from visual inspection of these figures or by correlating the position on the time scale of the wave reflection generated at the damage site with the laminate length, as was done in Sec. II for the narrow beam. Furthermore, the severity of damage could also be estimated, since the amplitude of the waves appearing in Figs. 13(b)

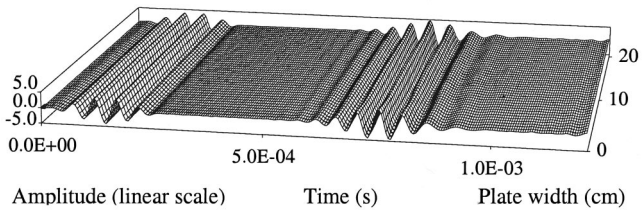


FIG. 12. The predicted in-plane displacement history of the nodes located along the AB boundary of the undamaged mesh.

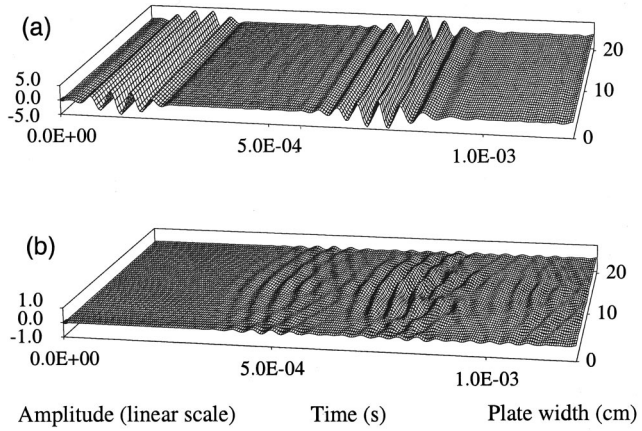


FIG. 13. The predicted in-plane displacement history of the nodes located along the AB boundary of the damaged mesh. (a) Time history for the 10-$\times$10-mm open-hole case; (b) arithmetic difference between undamaged and damaged response histories.

and 14(b) is in direct relation with the damage size; further study is needed to establish an explicit relationship. These results strongly suggest that damage could be found in large-area specimens using a permanently bonded linear array of piezoelectric sensor/actuator elements. This concept is currently under experimental evaluation.

IV. CONCLUDING REMARKS

The resonance spectrum method has been proposed as a technique for measuring low frequency, long wavelength flexural wave phase velocity. It is based on the study of spectra obtained by forced mechanical resonance of samples using sine-sweep excitation. The procedure can be used to obtain the dispersion curves of materials even when there is not enough information on its mechanical properties, or when its heterogeneous nature requires complex numerical techniques to model the systems. The dispersion curves were then used in the detection and characterization of damage.

The generation of Lamb waves using small and thin piezoceramic patches has been demonstrated and employed in the development of an on-line, structural integrity assess-

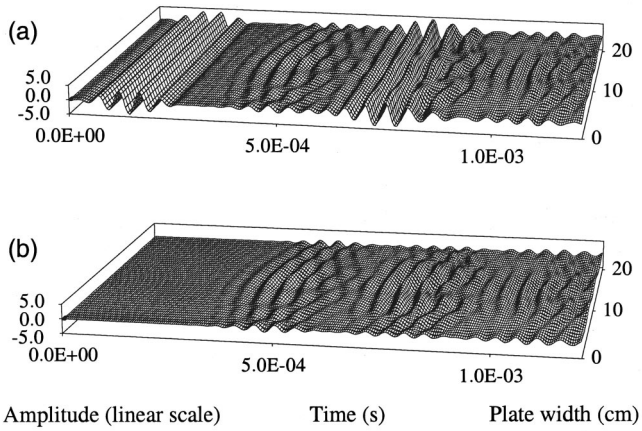


FIG. 14. The predicted in-plane displacement history of the nodes located along the AB boundary of the damaged mesh. (a) Time history for the 20-$\times$10-mm open-hole case; (b) arithmetic difference between undamaged and damaged response histories.

ment system capable of detecting delaminations in composite laminates. The study offers solution to some of the problems encountered with current Lamb wave generation techniques, where the use of standard ultrasonic probes render them unsuitable for the development of a built-in, health monitoring system due to their considerable size and shape. The inspection of CFRP laminates has been performed, producing promising results at excitation frequencies in the low ultrasonic range (<100 kHz) where IDTs have certain limitations to operate that restrict their use for the damage inspection of thick laminates.

Simple and effective signal processing has been employed for the detection of small changes in the structural response, whose minimal computational demand makes it appropriate for real-time continuous damage monitoring. Also, the good sensitivity to delaminations of relatively small size ($1\ \mathrm{cm}^2$), considerable propagation distances (over 2 m), and the very low requirement of electronic hardware offer a cost-effective solution for structural monitoring in terms of implementation and subsequent operation.

The finite-element analysis has shown the potential use of an array of transducers to achieve large-area scanning from fixed locations. It has been demonstrated that when the geometry of the array is chosen so that wave propagation in a single direction is induced, damage can be easily detected and located from the transducer's array response. The linear array chosen for the case of a plate is one example of this general concept. Future work will involve the optimization of the line array, whose number of sensors and spatial distribution will be determined from experimental and finite-element analyses, so that a proper inspection can be realized while economizing on hardware and processing complexity. Initially the piezoelectric elements will be individually bonded to the specimen; however, the concept has been conceived as to consist of a multielement array of ultrasonic transducers bonded to a thin flexible strip that can be permanently bonded to new and existing structures. The concept of embedding these devices into the laminate plate during the fabrication process will also be explored.

ACKNOWLEDGMENTS

This research was financially supported in part by the overseas research studentship (ORS) awards scheme, U.K., the National Council of Science and Technology (Conacyt), México, and the British Defense Evaluation and Research Agency (DERA), Farnborough. The authors wish to thank Mr. D. Hitchings from Imperial College and Professor P. T. Curtis from DERA for many helpful discussions.

[1] W.J. Percival and E.A. Birt, "A study of Lamb wave propagation in carbon-fiber composites," Insight **39**(10), 728–735 (1997).
[2] R.S.C. Monkhouse, P.D. Wilcox, and P. Cawley, "Flexible interdigital PVDF lamb wave transducers for the development of smart structures," Ultrasonics **35**(7), 489–498 (1997).
[3] P.D. Wilcox, P. Cawley, and M.J.S. Lowe, "Acoustic fields from PVDF interdigital transducers," IEE Proc.: Sci., Meas. Technol. **145**(5), 250–259 (1998).
[4] F.L. Degertekin and B.T. Khuri-Yakub, "Hertzian contact transducers for nondestructive evaluation," J. Acoust. Soc. Am. **99**, 299–308 (1996).
[5] R. Farlow and G. Hayward, "Real-time ultrasonic techniques suitable for implementing non-contact NDT systems employing piezoceramic composite transducers," Insight **36**(12), 926–935 (1994).
[6] S.G. Pierce, B. Culshaw, W.R. Philp, F. Lecuyer, and R. Farlow, "Broadband Lamb wave measurements in aluminum and carbon/glass fiber reinforced composite materials using non-contacting laser generation and detection," Ultrasonics **35**, 105–114 (1997).
[7] D. Alleyne and P. Cawley, "Optimization of Lamb wave inspection techniques," Nondestr. Test. Eval. **25**(1), 11–22 (1992).
[8] A. Birt, "Damage detection on carbon-fiber composites using ultrasonic Lamb waves," Insight **40**(5), 335–339 (1998).
[9] S.H. Díaz Valdés and C. Soutis, "Delamination detection in composite laminates from variations of their modal characteristics," J. Sound Vib. **228**(1), 1–9 (1999).
[10] S.H. Díaz Valdés, "Structural integrity monitoring of CFRP laminates using piezoelectric devices," Ph.D. thesis, Department of Aeronautics, Imperial College, September 2000.
[11] C.P. Mackinley, "Compressive failure of CFRP laminates containing pin-loaded holes," Ph.D. thesis, Imperial College, Department of Aeronautics, 2000.
[12] I.A. Viktorov, *Rayleigh and Lamb Waves—Physical Theory and Applications* (Plenum, New York, 1967).
[13] T. Pialucha, C.C.H. Guyott, and P. Cawley, "An amplitude spectrum method for the measurement of phase velocity," Ultrasonics **27**, 270–279 (1989).
[14] W. Sachse, and Y.H. Pao, "On determination of phase and group velocities of dispersive waves in solids," J. Appl. Phys. **49**(8), 4320–4327 (1978).
[15] R.G. Whit and R.J. Pinnington, "Practical application of the rapid frequency sweep technique for structural frequency response measurement," Aeronaut. J. **86**, 179–199 (1982).
[16] R.D. Blevins, *Formulas for Natural Frequency and Mode Shape* (Krieger, 1984).
[17] M.H. Datoo, *Mechanics of Fibrous Composites* (Elsevier Science, New York, 1991).
[18] B. Tangb, E.G. Henneke II, and R.C. Stiffler, "Low frequency flexural wave propagation in laminated composite plates," in *Proceedings of Acousto-Ultrasonics: Theory and Application*, edited by J.C. Duke, Jr. (Plenum, New York, 1988), pp. 45–65.
[19] N. Guo and P. Cawley, "Lamb wave reflection for quick nondestructive evaluation of large composite laminates," Mater. Eval. **52**(3), 404–411 (1994).
[20] D. Hitchings, "Finite Element Package FE77," Imperial College, Department of Aeronautics, 1997.

Experimental identification of finite cylindrical shell vibration modes

Lionel Haumesser, Dominique Décultot, Fernand Léon, and Gérard Maze
Laboratoire d'Acoustique Ultrasonore et d'Electronique, LAUE, UMR CNRS 6068, Université du Havre, Place Robert Schuman, 76610 Le Havre, France

(Received 3 July 2001; revised 5 February 2002; accepted 19 February 2002)

Acoustic scattering from a finite air-filled elastic cylindrical shell, immersed in water, is investigated. The shell is made of stainless steel and has a thickness to outer radius ratio of 17%. The considered dimensionless frequency range extends over $7 \leqslant \mathbf{k}_1 a \leqslant 22$ ($\mathbf{k}_1$: wave number in water, a: outer radius). Bistatic measurements are carried out to identify vibration modes related to the phase matching of the first guided wave, T_0, propagating on the shell. Both transducers, the emitter and the receiver, are positioned at the same angular distance with regard to the normal axis of the shell. The emitter transducer is fixed at a given position. In order to identify circumferential modes of vibration, the receiver transducer is made to rotate in the azimuthal plane, normal to the shell axis. Results obtained are plotted in functions of dimensionless frequency and azimuthal angle. Vibration modes along the shell's length are identified by moving the receiver transducer parallel to the shell axis. In this case, results are plotted in functions of dimensionless frequency and axial wave number. The experimental investigation is corroborated by theoretical results obtained from approximate calculations for thick finite cylindrical shells [Scot F. Morse *et al.*, J. Acoust. Soc. Am. **103**, 785–794 (1998)]. The evolution of the mode position with respect to the incidence angle is discussed so as to clarify peak patterns in backscattered resonance spectra. [DOI: 10.1121/1.1468877]

PACS numbers: 43.20.Ks, 43.40.Ey, 43.40.Fz, 43.20.Fn [JGM]

I. INTRODUCTION

In underwater acoustics, monostatic measurements have been extensively used, over the past years, in far-field detection of objects. In recent publications,[1,2] it has been shown that manmade objects differ from natural ones (e.g., rocks) in their elastic properties. In fact, manufactured objects, mines or pipes for instance, usually present geometrical and material characteristics that tend to enhance the propagation of surface waves and the phase matching. When interest is focused on scattering from finite targets, the monostatic approach is often used, such that surface waves radiate back in the direction of the emitter, after propagation through the extremities (reflections from truncations[3] or propagation in closed paths[4]).

Scattering enhancement from helical wave propagation on finite cylindrical shells presents useful features for target classification. Wave travel paths cover the whole object, so that propagation causes resonance phenomena proper to the shell as a whole, i.e., vibration modes relative to the target's circumferential and lengthwise dimensions. These phenomena depend on the aspect angle. Thus, measurements from successive incidence angles can be grouped together, in order to display the angular evolution of the backscattered pressure.

A number of authors consider the aspect dependence of finite cylindrical shells' acoustic signatures.[5–9] They investigate, in particular, the angular range within which resonant scattering involves contributions from mainly helical propagation of both whispering-gallery waves and guided waves.[8,10] These two types of waves are also known as, respectively, quasicompressional (the lowest order of which is denoted here as the S_0) and quasishear (T_0) helical waves.[9] Their propagation induces, at particular frequencies, a resonant state in the shell. It is reasonable to assume that minor damages in such objects (cracks in pipes, for instance) could affect their resonant response. Hence, the knowledge of mode organization within resonance spectra from an undamaged shell is necessary in ultrasonic nondestructive evaluation of structures, in general, and cylindrical ones in our case.

In this paper, a finite cylindrical shell is experimentally studied. First its backscattered resonance spectra, for incidence angles ranging from $\gamma_1 = -34°$ to $+34°$ (normal incidence to the shell axis is fixed at $\gamma_1 = 0°$) are presented. Then follows an investigation of vibration modes of resonance peaks related to the propagation of the T_0 wave. At given incidence angles, vibration modes in the circumference and along the length of the shell are successively identified. Experimental data thus obtained are systematically compared to results obtained from calculations for a finite thick shell using elasticity theory.[11]

II. THE TARGET AND THE EXPERIMENTAL SETUP

The investigated finite cylindrical shell is made of stainless steel of density $\rho = 7900\ \mathrm{kg\ m^{-3}}$, with longitudinal and transversal sound velocities of, respectively, $C_L = 5790\ \mathrm{m\ s^{-1}}$ and $C_T = 3100\ \mathrm{m\ s^{-1}}$. Dimensional characteristics of the object are: the outer radius $a = 6$ mm, the radii ratio $b/a = 0.83$ (b: inner radius), and the length $L = 65$ mm. The air-filled shell is closed at its extremities by rubber stoppers and immersed in a water tank (the sound velocity in water is $C_1 = 1470\ \mathrm{m\ s^{-1}}$). The stoppers are conical and truncated

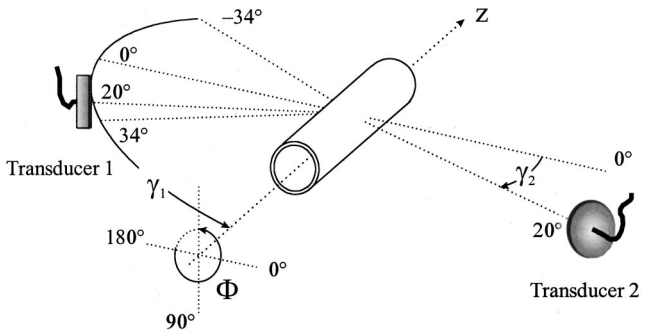


FIG. 1. Experimental arrangements; monostatic setup: transducer 1 used as both an emitter and a receiver; bistatic setup: transducer 1, fixed, used as an emitter and, transducer 2, moving around, angular position Φ (circumferential mode identification) or along the *z* direction (lengthwise mode identification), used as a receiver.

such that their elevation is equal to 4 mm and the diameter of the basement at half elevation is equal to *b*. The dimensionless frequency range extends over $7 \leqslant \mathbf{k}_1 a \leqslant 22$ ($\mathbf{k}_1$: wave number in water). In this range, the resonant behavior of the S_0 and the T_0 waves is observable.

The short-pulse monostatic and bistatic setups of the MIIR[12] (method of isolation and identification of resonances) are used to obtain experimental results presented in this paper (Fig. 1). Transducers are broadband, with central frequency f_c equal to 1.75 MHz ($\mathbf{k}_1 a = 45$) and the diameter of their radiating surfaces is 38 mm. A short electric pulse is converted into bulk acoustic wave by the emitter transducer. The time-domain responses of the insonified shell consist of echo waveforms made up of specular reflections and elastic wave reradiations. In the present investigation, resonance spectra are obtained from the FFT of time-domain responses, devoid of specular contributions. However, correction of the passbands of the transducers is not made (Fig. 2).

The monostatic part of the MIIR consists of the isolation of resonance frequencies.[13] Transducer 1 is used alternately as an emitter and a receiver to obtain the acoustic signature of the shell at given incidences. The distance between the transducer and the center of the object is 35 cm. Results are presented as resonance spectra, from which resonance frequencies can be read.

The bistatic part of the MIIR enables experimental determination of vibration modes. Both the emitter (transducer 1) and the receiver (transducer 2) are placed at equal angular distance with regard to the normal axis of the shell ($\gamma_1 = \gamma_2$; Fig. 1). The position of the emitter transducer remains fixed during identification procedures: (a) the circumferential vibration mode is labeled with an integer *n*; to obtain its identification, the receiver transducer is rotated in an azimuthal plane normal to the shell axis (angle Φ); the distances between the emitter transducer and the center of the object, and between the center of the object and the receiver transducer, are 45 and 35 cm, respectively; (b) the vibration mode in the shell's length is labeled with an integer *m*; its identification is performed by moving the receiver (transducer 2) parallel to the shell's axis (*z* axis); the distance between the object's axis and the line followed by the receiver transducer is 45 cm, and the distance between the emitter transducer and the center of the object is 35 cm.

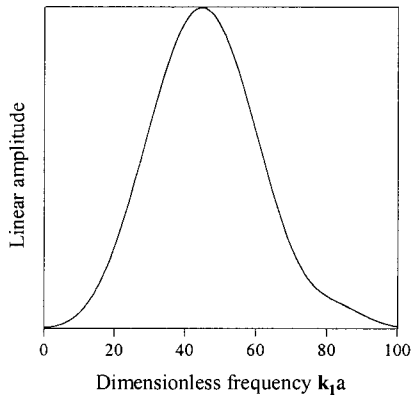


FIG. 2. Transducer passbands with central frequency: $f_C = 1.75$ Mhz ($\mathbf{k}_1 a = 45$).

III. MONOSTATIC MEASUREMENTS: ISOLATION OF RESONANCES

Surface waves generated at oblique incidence follow helical paths around the cylindrical shell, in the axial direction. In a previous work,[8] considering monostatic setup, echoes associated with the propagation of T_0 wave on finite cylindrical shells were found to be received into distinct groups. Successive echo groups were received after rising odd numbers of reflections at the extremities of the shell. It can be deduced from this that, in time-domain pulse responses, echo arrangements depend on both the radius and the length of the shell. In related spectra, a circumferential vibration mode appears consecutively through many peaks of resonance, contrary to the case of an infinite shell.

Experimental and calculated spectra are shown in the dimensionless frequency range $7 \leqslant \mathbf{k}_1 a \leqslant 22$ (Fig. 3). Experimental backscattered resonance spectra, from successive incidence angles (monostatic setup), are placed side by side to obtain Fig. 3(a). FFT is performed on time-domain responses, after the specular zone, over a duration of 450 μs. In the theoretical approach to the problem, the approximate calculations for finite cylindrical shell, developed by Morse *et al.*,[11] are used. The method combines full elasticity theory and the Kirchhoff diffraction integral; it is appropriate for thick-shell cases. The calculation of backscattered form function can be found in Appendix A of Ref. 11 (the reader is referred to Chap. 5 of Ref. 14 for more details). Figure 2 of Ref. 9 shows examples of form functions calculated in accordance with Ref. 11. In the present paper, two successive FFTs of the form function are performed to obtain Fig. 3(b): specular contributions are first removed from the form function. Shell parameters given in Sec. II are used in the computations. Resonances associated with the propagation of the S_0 wave are observable between the normal incidence and the angle given by the velocity at low frequencies of the first longitudinal mode for plates (γ_1 from 0° to ±16.3°). Resonance peaks of the T_0 wave are observable from normal incidence to the critical angle given by C_T, the transversal sound velocity (γ_1 from 0° to ±28.3°). The refraction effect shown by Conoir *et al.*,[15] in the case of an infinite cylinder, is once again observed here: resonances due to the phase matching of circumnavigating helical surface waves stand on

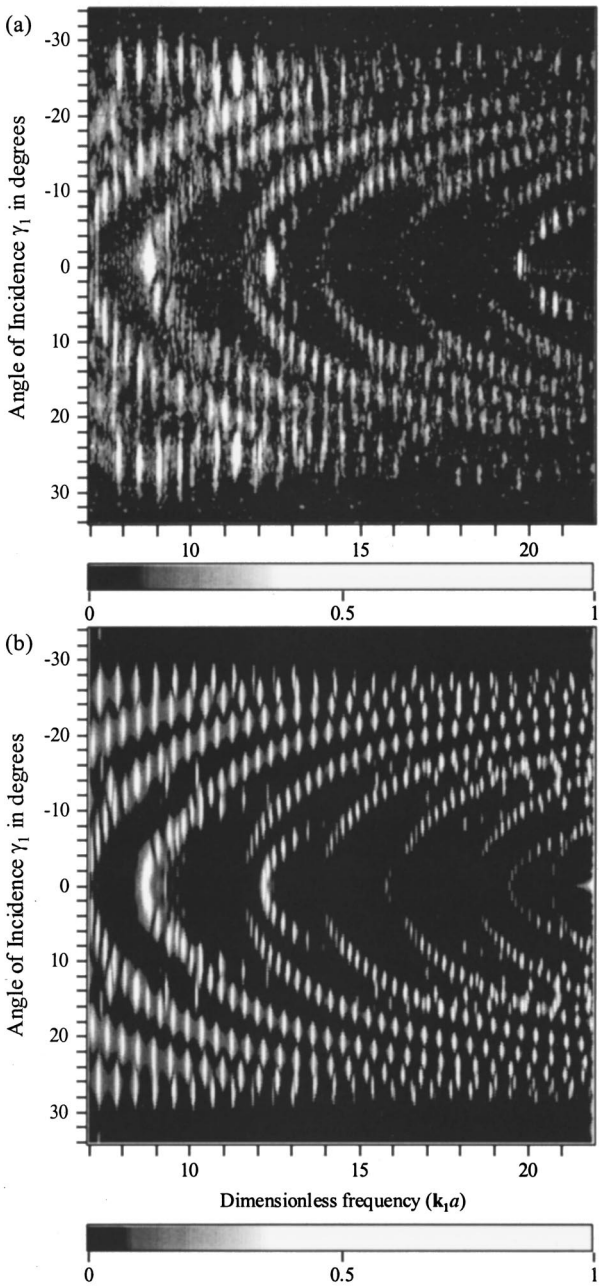


FIG. 3. (a) Experimental and (b) theoretical backscattered resonance spectra from an immersed empty finite cylindrical shell ($b/a=0.83$ and $L/a=10.83$), shown as a function of incidence angle γ_1 from $-34°$ to $+34°$, at half-degree increments. Linear amplitudes are normalized with respect to the peak value in each figure ($\gamma_1=0°$ and $\mathbf{k}_1 a=12.2$ in both cases).

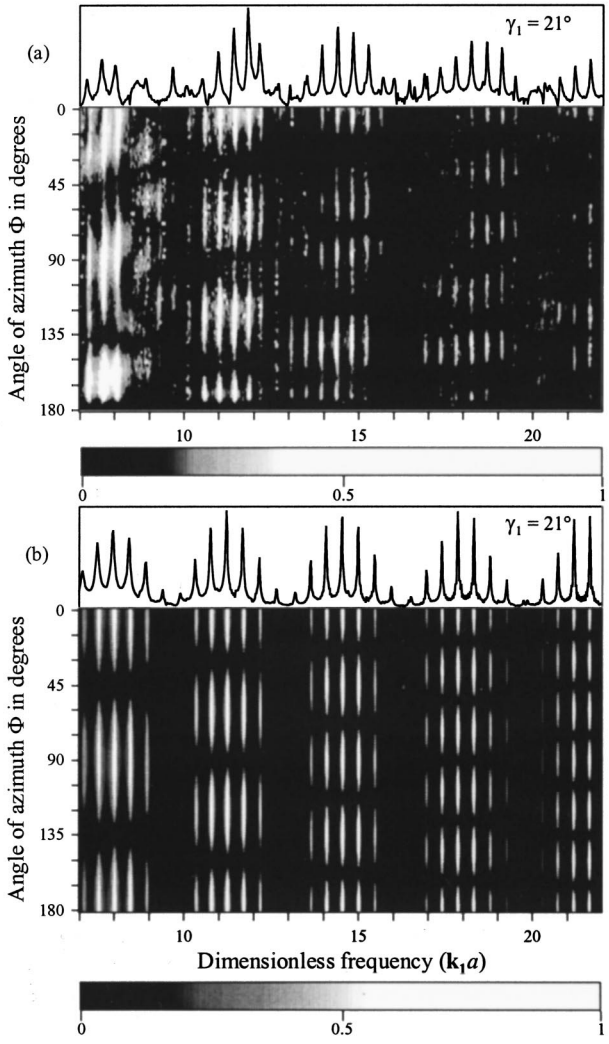


FIG. 4. Planes of modal identification of circumferential vibration mode n at $\gamma_1=\gamma_2=21°$. Transducer 2 rotates in an azimuthal plane normal to the shell's main axis, with azimuthal angle Φ at 2° increments. (a) Experimental results; linear amplitudes are normalized with respect to the peak value at $\Phi=166°$ and $\mathbf{k}_1=7.7$; the upper part shows the relevant backscattered resonance spectrum. (b) Theoretical results; linear amplitudes are normalized with respect to the peak value at $\Phi=180°$ and $\mathbf{k}_1 a=12.2$; relevant spectrum (upper part) is extracted from Fig. 3(b).

trajectories which shift towards high frequencies for incidence angles deviating away from normal incidence. In the case of a finite cylindrical shell, several peaks of resonances in a spectrum can be associated to the same circumferential vibration mode of the T_0 wave.[16] These peaks differ by their vibration modes m along the shell's length.

In the following discussion, an attempt is made to clarify the organization of vibration modes within resonance spectra. Sections IV A and IV B, respectively, deal with the identification of vibration modes in the circumference and along the length of the shell.

IV. BISTATIC MEASUREMENTS: IDENTIFICATION OF RESONANCES

A. Circumferential vibration modes

The identification of circumferential modes of vibration is carried out at $\gamma_1=\gamma_2=21°$. The receiver (transducer 2) rotates around the shell, in an azimuthal plane normal to the shell axis, the angle Φ varying from 0° to 180°, at 2° intervals. Results are plotted in functions of the dimensionless frequency and the azimuthal angle. The presentation is herein termed *plane of modal identification*. Experimental spectra are shown in Fig. 4(a). The lower part yields the mode identification. At a given resonance frequency, the circumferential vibration mode (n) is the number of pressure lobes established on the semicircumference. The upper part of Fig. 4(a) presents the resonance spectrum from backscattering with transducer 1, obtained in the same conditions as those in Fig. 3(a). Results of calculations performed using

TABLE I. Resonance frequencies and associated identified circumferential vibration modes of the T_0 wave on finite cylindrical shell at $\gamma_1 = 21°$. The experimental/theoretical correspondence is done for the closest frequencies.

Incidence angle $\gamma_1 = 21°$			
Experimental		Theoretical	
Mode n	Frequency $\mathbf{k}_1 a$	Mode n	Frequency $\mathbf{k}_{1a}$
2	7.3	2	7.1
2	7.7	2	7.5
2	8.1	2	8.0
⋯	⋯	2	8.5
⋯	⋯	2	8.9
⋯	⋯	3	9.4
3	10.2	3	9.9
3	10.6	3	10.3
3	11.0	3	10.8
3	11.4	3	11.2
3	11.9	3	11.7
3	12.2	3	12.2
⋯	⋯	3	12.7
⋯	⋯	4	13.2
⋯	13.5	4	13.7
4	14.0	4	14.1
4	14.4	4	14.5
4	14.9	4	15.0
4	15.3	4	15.5
4	15.7	⋯	⋯
⋯	⋯	5	16.5
⋯	16.9	5	17.0
5	17.4	5	17.4
5	17.8	5	17.9
5	18.3	5	18.3
5	18.7	5	18.8
5	19.1	5	19.3
5	19.5	⋯	⋯
⋯	⋯	6	20.3
⋯	20.7	6	20.7
6	21.2	6	21.2
6	21.6	6	21.7

the same parameters and similar angular reception conditions are shown in Fig. 4(b). This spectrum is an extract from Fig. 3(b). Resonance frequencies and identified circumferential mode values are read from Figs. 4(a) and (b) and are listed in Table I. On both experimental and theoretical planes of modal identification, the same mode of circumferential vibration is identified for several distinct peaks of resonances. This observation is illustrated in Fig. 5, through 360°. Experimentally, mode $n=3$ is established at, among other frequencies, $\mathbf{k}_1 a = 11.0$ (a), and $\mathbf{k}_1 a = 11.9$ (b). Open lobes in the figure are due to the passage of the receiver between the emitter and the shell (at $\Phi = 180°$). The theoretical approach confirms experimental identification (c); other calculated pressure patterns in the azimuthal plane for mode $n=3$, obtained at frequencies listed in Table I, are similar, except for their amplitude. In the next section, the identification is supplemented by the determination of vibration mode m, along the length of the shell.

B. Lengthwise vibration modes

The identification of vibration modes along the length of the shell is carried out at $\gamma_1 = \gamma_2 = 20°$. The shell is linearly

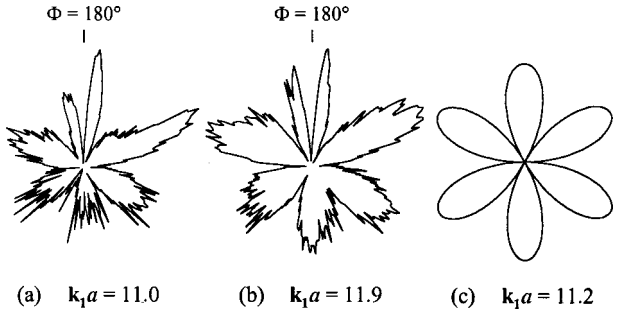


FIG. 5. Illustration of circumferential vibration mode splitting for $n=3$, from 1° increment statements, over a range of azimuthal angle Φ from 0° to 360°. Linear amplitude is not normalized. (a,b) Experimental results at $\mathbf{k}_1 a = 11.0$ and 11.9, respectively; at $\Phi = 180°$, transducer 2 stands between the emitter and the shell such that no signal is received; (c) theoretical result at $\mathbf{k}_1 a = 11.2$.

scanned along its axis (z axis) by the receiver (transducer 2), at 0.2-mm increments, over 52 mm of the total length of the shell. From the spectra corresponding to z-axis positions, spatial (axial) Fourier transform is performed over the scanned length of the shell. The relevant wave number ($k_m = m\pi/L$) is linked to the vibration mode m when an integer number of half wavelengths is established along the length of the shell. Results are plotted as a function of dimensionless frequency and vibration mode m. Such a presentation provides the experimental identification of m [lower part of Fig. 6(a)]. The upper part of Fig. 6(a) is the backscattered resonance spectrum obtained with transducer 1; it is obtained in the same conditions as those in Fig. 3(a). The calculation of the theoretical backscattered resonance spectrum presented in the upper part of Fig. 6(b) [and extracted from Fig. 3(b)] takes into account two mode summations, in the circumference and along the length of the shell. The frequency/vibration mode presentation [lower part of Fig. 6(b)] is obtained by ''cascading'' mode m by mode m of the latter spectrum, i.e., one line of this representation is the contribution of circumferential modes for one mode m along the length. Here, it would be more accurate to call it *plane of modal isolation*. This presentation is preferred to that adopted in Sec. IV A, since it enables theory/experiment confrontation. However, because of numerical constraints in the calculations,[9] it is obtainable from grouping three zones: the first covers $7 \leqslant \mathbf{k}_1 a \leqslant 22$ and $5 \leqslant m \leqslant 22$, the second covers $14.5 \leqslant \mathbf{k}_1 a \leqslant 22$ and $23 \leqslant m \leqslant 30$; values of the third zone are arbitrarily set to zero. The small white areas visible for $m=26$ at $\mathbf{k}_1 a = 14$ and $\mathbf{k}_1 a = 22$ [Fig. 6(b)] are due to the rectangular window effect from the FFT computation of the second zone. In the frequency/vibration mode planes, the resonances stand on successive oblique trajectories, each corresponding to one circumferential vibration mode. Hence, the presentation enables the arrangement of resonances in the circumferential mode n, but does not lead to the determination of this mode. The identification along the length is illustrated in Fig. 7, through extracts from Figs. 6(a) and (b), for $m=15$ and 20. Mode broadening in the experimental data is introduced by the number of measurements realized during reception along the shell axis, and the ''Hamming''-type filter used to minimize secondary lobes due to spatial Fourier

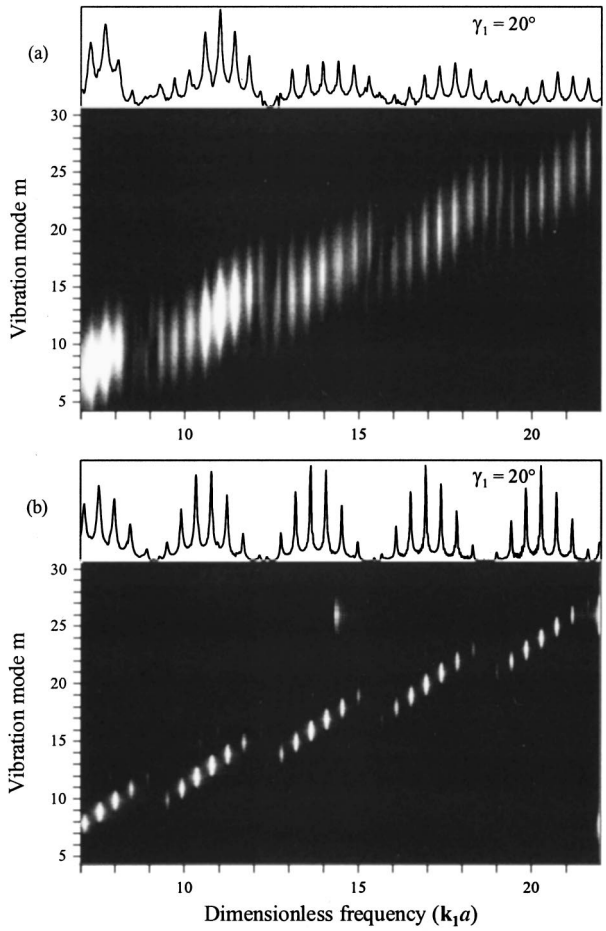


FIG. 6. Planes of modal identification of lengthwise vibration mode m at $\gamma_1 = \gamma_2 = 20°$. Amplitude is not normalized and increases linearly from the white to the black zone. (a) Experimental results; the upper part is the corresponding backscattered resonance spectra. Transducer 2 scans the shell's z axis over 52 mm of the total length, at 0.2-mm increments. (b) Theoretical results; relevant spectrum (upper part) is extracted from Fig. 3(b).

transform. Results from identification in Figs. 6(a) and (b) are summarized in Table II. For some given mode m, an average frequency shift equal to 0.4 is observed between theoretical and experimental values. This difference may be caused by the use of characteristics of stainless steel in the computations (C_L, C_T, and ρ; see Sec. II).

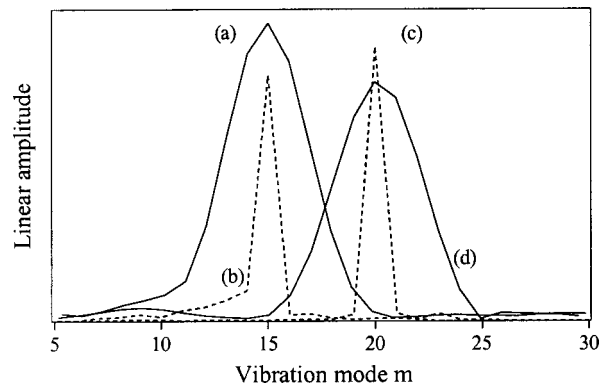


FIG. 7. Experimental (solid lines) and theoretical (dashed lines; value points have been connected for clarity) identification of vibration modes $m = 15$ and $m = 20$ in the shell length, at (a) $\mathbf{k}_1 a = 11.9$; (b) $\mathbf{k}_1 a = 11.7$; (c) $\mathbf{k}_1 a = 17.0$, and (d) $\mathbf{k}_1 a = 17.4$, all extracted from Figs. 6(a) and (b).

TABLE II. Resonance frequencies and associated identified circumferential and lengthwise vibration modes of the T_0 wave on finite cylindrical shell at $\gamma_1 = 20°$.

Incidence angle $\gamma_1 = 20°$				
Experimental		Theoretical		
Mode	Frequency	Mode		Frequency
m	$\mathbf{k}_1 a$	m	n	$\mathbf{k}_1 a$
8	7.3	8	2	7.1
9	7.7	9	2	7.5
10	8.1	10	2	8.0
⋯	8.5	11	2	8.5
⋯	⋯	12	2	8.9
9	9.3	⋯	⋯	⋯
10	9.7	10	3	9.4
11	10.2	11	3	9.9
12	10.6	12	3	10.3
13	11.0	13	3	10.8
14	11.4	14	3	11.2
15	11.9	15	3	11.7
16	12.2	⋯	⋯	⋯
14	13.1	14	4	12.8
15	13.5	15	4	13.2
16	14.0	16	4	13.7
17	14.4	17	4	14.1
18	14.9	18	4	14.5
19	15.3	19	4	15.0
17	16.1	⋯	⋯	⋯
18	16.5	18	5	16.1
19	16.9	19	5	16.5
20	17.4	20	5	17.0
21	17.8	21	5	17.4
22	18.3	22	5	17.9
23	18.7	23	5	18.3
⋯	19.1	⋯	⋯	⋯
⋯	19.5	⋯	⋯	⋯
22	19.9	22	6	19.4
23	20.3	23	6	19.9
24	20.7	24	6	20.3
25	21.2	25	6	20.7
26	21.6	26	6	21.2
⋯	⋯	27	6	21.7

In Table II, significant resonance peaks are theoretically identified by their vibration modes (n,m). Among them, many have the same vibration mode m. These resonances, however, do not have the same circumferential vibration mode n. This particularity is highlighted when only one mode along the length is considered, taking into account the circumferential modal contributions: mode $m = 18$ in Fig. 8. Further, in this same figure, the frequency position of resonance peaks with given vibration modes remains the same for different incidence angles: it is the case for the resonance $n = 4$ and $m = 18$ at $\mathbf{k}_1 a = 14.5$. This example shows complete identification of the vibration modes (n,m) of the same resonance for a number of closely spaced incidence angles. Hence, several resonances are found at identical frequency positions for $\gamma_1 = 21°$ and $\gamma_1 = 20°$ (Tables I and II).

V. CONCLUSION

Bistatic measurements of acoustic scattering from a finite air-filled elastic cylindrical shell, immersed in water, are considered. Resonances relevant to the T_0 wave are investigated. Vibration modes in characteristic dimensions of the

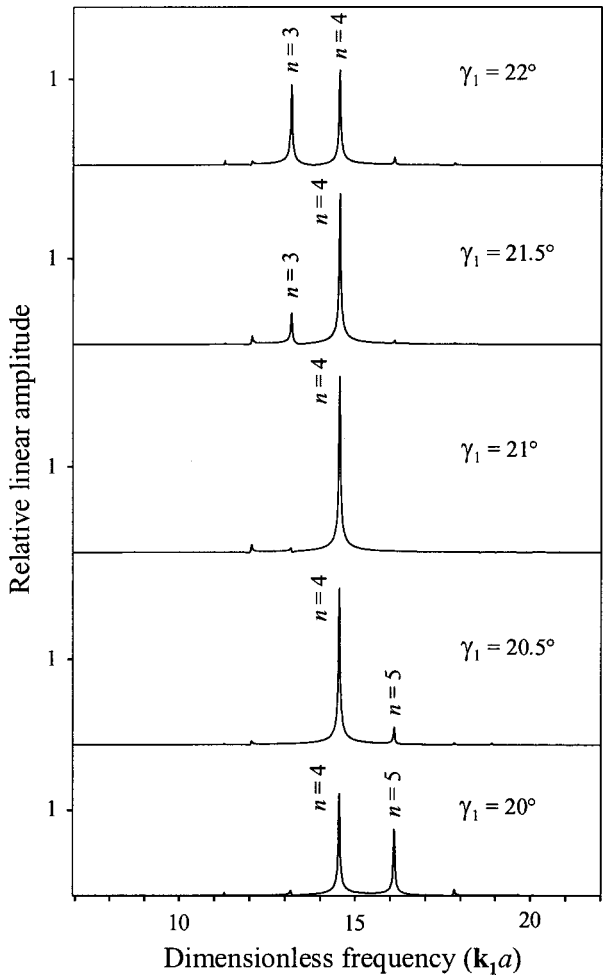


FIG. 8. Theoretical resonance spectra for $m=18$, at successive incidence angles γ_1.

shell (i.e., circumference and length) are experimentally identified. Experimental results are compared to approximate calculations for the thick shell. The identification performed in an azimuthal plane normal to the shell's axis shows the possibility, for several resonances within the spectrum, to have the same mode of circumferential vibration *n*. The second identification, performed along the shell axis, provides the determination of lengthwise vibration mode *m*. Resonances can be fully identified experimentally by their vibration modes (n,m). This enables experimental differentiation of resonances whose circumferential vibration mode is the same. Within the spectrum, one mode *m* can be obtained for resonances with various modes *n*. Further, a resonance of given vibration modes is observable at a number of successive angles at the same frequency, in backscattered spectra.

ACKNOWLEDGMENT

The authors wish to thank Dr. J. G. Chiumia for his contribution to the readability of this paper.

[1] B. Zerr, A. Tesei, A. Maguer, B. H. Houston, and P. A. Sletner, "Classification of underwater elastic objects based on aspect dependence of their acoustic signature," 5th European Conference on Underwater Acoustics, Lyon, France, Conference Proceedings, 875–880 (2000).

[2] M. Tran Van Nhieu, M. Gensane, S. Fioravanti, A. Tesei, A. Maguer, B. Woodward, and P. A. Lepper, "Detection of a buried water-filled cylindrical shell by the wavelet transform technique," 5th European Conference on Underwater Acoustics, Lyon, France, Conference Proceedings, 1091–1096 (2000).

[3] K. Gipson and P. L. Marston, "Backscattering enhancements due to reflection of meridional leaky Rayleigh waves at the blunt truncation of a tilted solid cylinder in water: Observations and theory," J. Acoust. Soc. Am. **106**, 1673–1680 (1999).

[4] N. Touraine, L. Haumesser, D. Décultot, G. Maze, A. Klauson, and J. Metsaveer, "Analysis of the acoustic scattering at variable incidences from an extra thin cylindrical shell bounded by hemispherical endcaps," J. Acoust. Soc. Am. **108**, 2187–2196 (2000).

[5] M. L. Rumerman, "Contribution of membrane wave reradiation to scattering from finite cylindrical steel shells in water," J. Acoust. Soc. Am. **93**, 55–65 (1993).

[6] X.-L. Bao, "Echoes and helical surface waves on finite elastic cylinder excited by sound pulses in water," J. Acoust. Soc. Am. **94**, 1461–1466 (1993).

[7] M. Tran Van Nhieu, M. Gensane, A. Tesei, S. Fiovaranti, and A. Maguer, "Scattering by a fluid-filled finite cylindrical shell: Comparison theory-experiments," 4th European Conference on Underwater Acoustics, Rome, Italy, Conference Proceedings, 819–824 (1998).

[8] L. Haumesser, A. Baillard, D. Décultot, and G. Maze, "Behavior of first guided wave on finite cylindrical shells of various lengths: Experimental investigation," J. Acoust. Soc. Am. **109**, 583–590 (2001).

[9] S. F. Morse and P. L. Marston, "Meridional ray contributions to scattering by tilted cylindrical shells above the coincidence frequency: Ray theory and computations," J. Acoust. Soc. Am. **106**, 2595–2600 (1999).

[10] F. Léon, F. Lecroq, D. Décultot, and G. Maze, "Scattering of an obliquely incident acoustic wave by an infinite hollow cylindrical shell," J. Acoust. Soc. Am. **91**, 1388–1397 (1992).

[11] S. F. Morse, P. L. Marston, and G. Kaduchak, "High-frequency backscattering enhancements by thick finite cylindrical shells in water at oblique incidence: Experiments, interpretation, and calculations," J. Acoust. Soc. Am. **103**, 785–794 (1998).

[12] P. Pareige, P. Rembert, J. L. Izbicki, G. Maze, and J. Ripoche, "Méthode impulsionnelle numérisée (MIN) pour l'isolement et l'identification des résonances de tubes immergés," "Digitized pulse method (MIN in French) for isolation and identification of resonances of fluid-loaded cylindrical shells," Phys. Lett. A **135**, 143–146 (1989).

[13] G. Maze and J. Ripoche, "Méthode d'Isolement et d'Identification des Résonances (M.I.I.R.) de cylindres et de tubes soumis à une onde acoustique plane dans l'eau," "Method of isolation and identification of resonances (M.I.I.R.) of cylinders and cylindrical shells insonified by a plane acoustic wave in water," Rev. Phys. Appl. **18**, 319–326 (1983).

[14] S. F. Morse, "High Frequency Acoustic Backscattering Enhancements for Finite Cylindrical Shells in Water at Oblique Incidence," Ph.D. dissertation, Washington State University, Department of Physics, 1998.

[15] J. M. Conoir, P. Rembert, O. Lenoir, and J. L. Izbicki, "Relation between surface helical waves and elastic cylinder resonances," J. Acoust. Soc. Am. **93**, 1300–1308 (1993).

[16] F. Lecroq, G. Maze, D. Décultot, and J. Ripoche, "Acoustic scattering from an air-filled cylindrical shell with welded flat plate endcaps: Experimental and theoretical study," J. Acoust. Soc. Am. **95**, 762–769 (1994).

Phased array element shapes for suppressing grating lobes

F. Joseph Pompei[a] and Shi-Chang Wooh[b]
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 7 August 2001; revised 28 November 2001; accepted 17 December 2001)

Most techniques for suppressing grating lobes in phased arrays while relaxing the interelement spacing requirement involve redistributing array elements in sparse aperiodic patterns, or varying the transmit-receive beam patterns. An alternative is presented which uses oversized array elements to eliminate grating lobes as a direct consequence of the element shape. It is shown that by using carefully shaped, overlapping elements, maximum scan angle can be exchanged for a reduced interelement spacing requirement. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1460924]

PACS numbers: 43.20.Rz, 43.20.Bi [DEC]

I. INTRODUCTION

Phased arrays have long been used for both transmission and reception of sound waves in a variety of acoustic applications. Ultrasonics, in particular, has a long history of phased arrays for applications such as underwater acoustics, medical imaging, ultrasonic therapy, and nondestructive evaluation (NDE).

The most important parameters affecting the cost and performance of a phased array system are the number of elements and the interelement spacing necessary to provide a desired steering response. In a traditional periodic array, an interelement spacing of less than half the wavelength ($\lambda/2$) is required to mitigate detrimental grating lobes.[1,2] Because the main lobe width is dependent only on the spatial extent of the array, the generation of a narrow beam will usually require a large array and an inordinate number of individually driven elements.

Various methods have been proposed to relax the interelement spacing requirement to create sparse arrays of fewer elements with reduced grating lobes. Because the grating lobes are a result of the periodicity of the element positions, they can be reduced through the use of a random or aperiodic distribution of elements, although at the expense of a reduced dynamic range.[1,3,4] Others have proposed using different element patterns for transmit versus receive modes,[5–8] or by relying on very short pulses.[9] While these techniques are generally successful, they are applicable only for pulse-echo systems, not continuous wave systems, such as loudspeakers, microphone arrays, or parametric audio sources.[10]

Most of these earlier studies involve addressing the position and distribution of the individual elements, but little attention has been paid to the characteristics of the elements themselves. It is well known[11] that the directivity envelope pattern is wholly determined by the shape and size of the constituent elements, so it is logical that the element shape can be used to control various steering characteristics. In particular, elements larger than the interelement spacing will narrow the directivity envelope substantially, naturally suppressing grating lobes for a limited steering angle. Aspects of this technique have been explored for antenna arrays,[12] but it has received little attention in the context of acoustics or ultrasonics.

In this study, it is shown that overlapping arrays of various oversized elements can be used to suppress grating lobes, and allow a relaxed interelement spacing requirement for a limited steering angle. A brief review of relevant phased array theory is introduced, followed by a consideration of overlapped arrays of common element shapes which suppresses grating lobes to some extent, at the expense of a nonuniform main lobe amplitude. An optimal element shape is then developed, which effectively squelches grating lobes without influencing the main lobe. For convenience, only one-dimensional arrays are considered in this paper, although most of the techniques may also be applicable to higher-dimensional arrays.

II. BASIC THEORY

The far-field response of an arbitrary source condition can be derived using Huygen's principle, which may then be tailored to describe the specific case of a traditional linear array with uniformly spaced elements. This is largely based on the analysis given in Refs. 1, 2, and 11.

A. Huygen's principle

Huygen's principle states that any wave-producing source can be modeled as an infinite number of individual sources distributed identically to the original source. A pulsating point source radiates energy in spherical waves described by the equation

$$p_{\text{point}}(R) = \frac{dp_0}{R} e^{j(\omega t - kR)}, \tag{1}$$

where dp_0 is the source amplitude, R is the distance from the source to the point of interest, ω is the driving frequency, k is the wave number, and j is the unit imaginary number.

With an infinite baffle assumption, modeling a one-dimensional source with arbitrary source amplitude distribution $s(x)$, as shown in Fig. 1, results in the pressure distribution in an integral form:

$$p(r,\theta) = e^{j\omega t} \int \frac{s(x)}{R} e^{-jkR}\, dx. \tag{2}$$

[a]Media Laboratory, Massachusetts Institute of Technology, Room E15-492, 20 Ames St., Cambridge, MA 02139. Electronic mail: pompei@media.mit.edu
[b]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Room 1-272, 77 Massachusetts Avenue, Cambridge, MA 02139. Electronic mail: scwooh@mit.edu

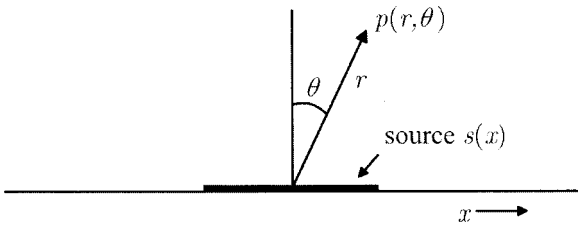


FIG. 1. The geometry of the one-dimensional source under consideration.

The source amplitude $s(x)$ is the contribution of source element dx to the pressure at point p, and relates to the source's normal vibration velocity amplitude $u(x)$ as $s(x) = (\rho_0 \omega/2\pi)u(x)$.

The $e^{j\omega t}$ term may be omitted for a linear and nondispersive system, and R can be approximated as

$$R = (r^2 + x^2 - 2rx \sin\theta)^{1/2} \approx r - x \sin\theta + \frac{x^2}{2r}. \tag{3}$$

Substituting this approximation into the integral gives

$$p(r,\theta) = \frac{1}{r} e^{-jkr} \int s(x) e^{jkx \sin\theta} e^{-jk(x^2/2r)}\, dx. \tag{4}$$

The last factor in the integral is significant only for small r, when $r \ll 1/2kx^2$, the Fresnel distance. The other factors are not dependent on r, and therefore it describes the pressure in the *far field* of the source. For simplicity, this analysis will address only far-field effects, limiting the extent of the array to those dimensions for which this approximation is satisfied.

B. Fourier analysis

If the change in variables $x' = x/\lambda$ and $\beta = \sin\theta$ is made while disregarding the amplitude scaling with r, the integral describing the far-field directivity becomes

$$H(\beta) = \int s(x') e^{j2\pi x' \beta}\, dx'. \tag{5}$$

This is the Fourier transform of the source distribution function scaled by $1/\lambda$. This important result makes the analysis of the far field straightforward and intuitive.

As an alternative, which is more suitable for discrete Fourier transform (DFT) simulation, one can substitute $\Omega = \beta/\lambda$, and arrive at

$$H\left(\frac{\beta}{\lambda}\right) = f(\Omega) = \int s(x) e^{j2\pi\Omega x}\, dx. \tag{6}$$

In this case, the scaling with λ is taken *after* the Fourier transform, so that downsampling prior to taking the DFT is not necessary.

Simulation Steps:

(1) Given the source distribution $s[n]$, where $n = xT$ is a sufficiently fine spatial sampling of the source function
(2) Zero pad $s[n]$
(3) Compute the DFT; $H(\Omega) = \mathrm{DFT}\{s[n]\}$
(4) Rescale the angle axis $\beta = \lambda\Omega$
(5) Normalize (if desired).

C. Steering

Beam steering is simply shifting the overall response with respect to β, such that

$$H'(\beta) = H(\beta + \beta_s), \tag{7}$$

where β_s is the steering angle. This constant shift in the "transform" domain is equivalent to multiplying by $e^{jk\beta_s x}$ in the "source" domain, i.e.,

$$s'(x) = s(x) e^{jk\beta_s x}. \tag{8}$$

This factor is simply a space-dependent phase shift, or time delay, distributed linearly across the array.

D. Polar cut

When analyzing the one-dimensional directivity of a two-dimensional source, it is useful to reduce the source function to a one-dimensional equivalent, as described in Ref. 1. Under a similar far-field approximation as discussed earlier (where the distance from the array is much larger than the array itself), two-dimensional source $\hat{s}(x,y)$ can be reduced to a one-dimensional equivalent:

$$s(x) \simeq \int \hat{s}(x,y)\, dy. \tag{9}$$

For rectangular elements in a traditional linear array, this integral does not vary over x, so this technique offers little advantage. This method becomes very useful, however, when shaped elements are described in the next section.

E. Discrete array

A source $s(x)$ comprised of a discrete array of identical elements, shown in Fig. 2(a), can be treated as an infinite number of copies of the element source function $s_0(x)$ multiplied by an overall aperture function $w(x)$, which is generally rectangular:

$$s(x) = s_0(x) * \left(w(x) \cdot \sum_{n=-\infty}^{\infty} \delta(x - x_n) \right), \tag{10}$$

where x_n is the location of the nth element, and the asterisk denotes convolution. For a periodic array, where $x_n = nd$, the far-field response $H(\beta)$, given by the Fourier transform of the source function $s(x)$ is

$$H(\beta) = S_0(\beta) \cdot \left(W(\beta) * \sum_{m=-\infty}^{\infty} \delta\left(\beta - \frac{m\lambda}{d} \right) \right). \tag{11}$$

This can be conveniently written as the product of two directivity functions, $H_1(\beta)$ and $H_2(\beta)$:

$$H(\beta) = H_1(\beta) H_2(\beta), \tag{12}$$

with $H_1(\beta) = S_0(\beta)$ and $H_2(\beta)$ equal to the terms in parentheses. In this study, only the quantity $H_1(\beta)$ is of interest, as it is wholly dependent on the individual element characteristics.

An example array with rectangular elements is shown in Fig. 2(a). Here, the rectangular elements have width a and the overall array has a width of D. The normalized responses from these components, plotted in Fig. 2(b) are

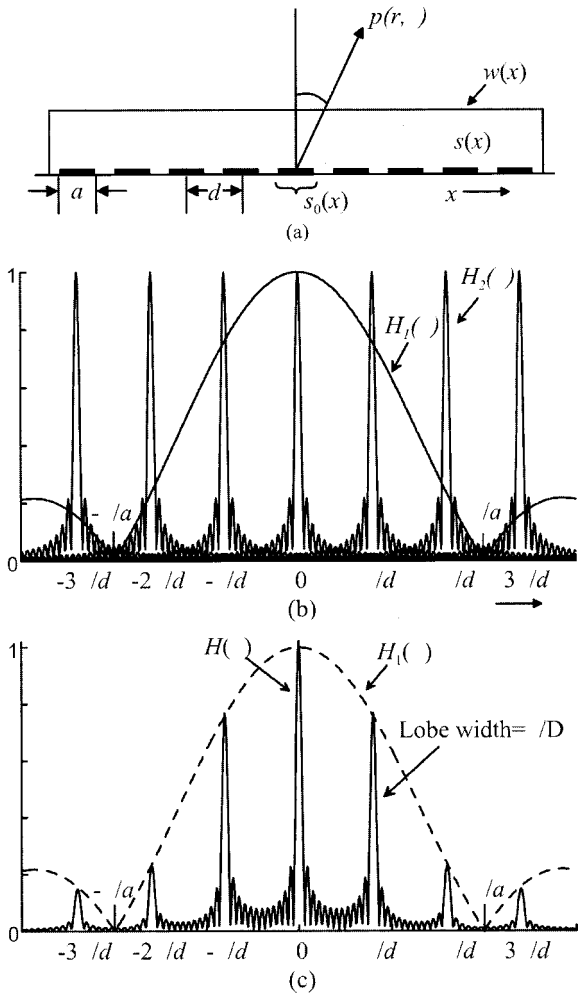


FIG. 2. A typical linear array, with rectangular elements. The physical components of the array, including the elements of width a, interelement spacing d, and overall array size D, are shown in (a). The contributions of the directivity function due to the element size and overall array size are shown in (b). The final directivity function $H(\beta)$ is shown in (c). Recall that $\beta=\sin\theta$.

$$s_0(x)=\mathrm{rect}(a)\Leftrightarrow H_1(\beta)=\mathrm{sinc}(\pi a\beta/\lambda), \tag{13}$$

$$w(x)=\mathrm{rect}('D)\Leftrightarrow W(\beta)=\mathrm{sinc}(\pi D\beta/\lambda). \tag{14}$$

The rectangular function $\mathrm{rect}(a)$ is unity for $-a/2\leqslant x\leqslant a/2$ and zero otherwise, and $\mathrm{sinc}(x)=\sin(x)/x$. The constant interelement spacing of d creates periodic lobes at intervals d/λ in $H_2(\beta)$. The final response $H(\beta)$ is shown in Fig. 2(c). Only the interval $-1\leqslant\beta\leqslant 1$ maps to physical space, so this interval is usually termed the *visible region*. Any additional lobes present in this region are termed *grating lobes*.

This result illustrates two important intuitive points:

(i) The far-field response consists of an infinite set of copies of the overall aperture response $W(\beta)$, spaced by λ/d.

(ii) This total response is then modulated by an *element* response, or the directivity envelope, $H_1(\beta)$.

Thus, in the ''source'' domain, the individual element response influences the overall response of the array in the transform domain, while the overall window in the source domain affects the individual elements in the transform domain.

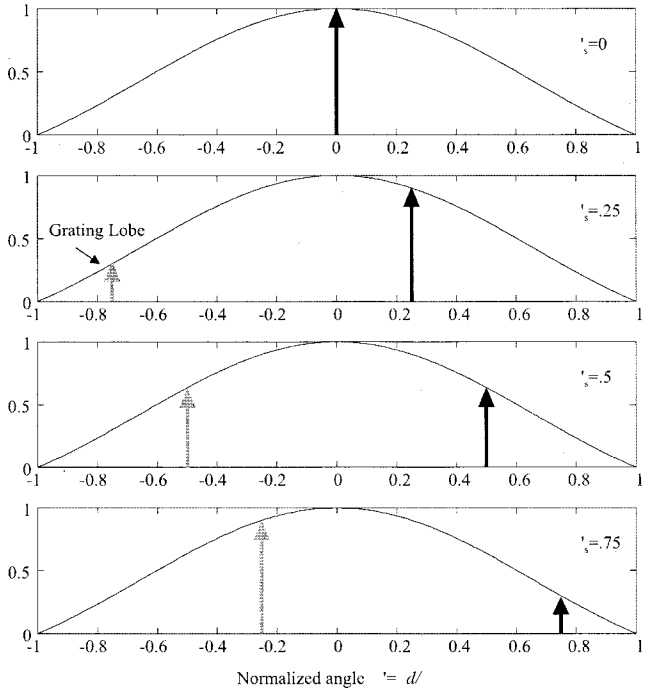


FIG. 3. The directivity of a dense array with element size a equal to the element spacing d. The plots show the changing position and amplitude of the main lobe and grating lobe as the beam is steered. Notice that the magnitude of the grating lobe is equal to that of the main lobe when $\beta_s=\lambda/2a$.

F. Dense arrays

In many applications, it is desirable to maximize sensitivity per unit length of the array, which necessitates the maximization of the density of the array. In a traditional linear array, this is accomplished by making the interelement spacing d equal as close as possible to the element size a. This type of array may be called a *dense array*.

For an array with element size a equal to its interelement spacing d, the grating lobes in an unsteered beam are coincident with zeros in the directivity envelope $H_1(\beta)$, as shown in the upper plot of Fig. 3.

For clarity, the magnitudes are plotted with respect to a normalized angle $\beta'=\beta d/\lambda$, so that variations in wavelength will simply rescale the horizontal axis, and affect the visible region, not the shape of the response.

As the beam is steered, the grating lobe increases in magnitude, while the main lobe decreases. The main lobe and grating lobe have equal amplitudes when $\beta_s'=1/2$, or equivalently when $\beta_s=\lambda/2a$. This sets a limit on useful steering angle to a small range of angles where $\beta_s<\lambda/2a$.

G. Steering performance

The most important indicators of phased array performance which depend on array geometry are (1) the main lobe magnitude M, (2) the grating lobe magnitude G, and (3) the ratio between the amplitude of the main lobe versus that of the grating lobe $\zeta=M/G$. Recall that the beam width is independent of the element shape, and is simply associated

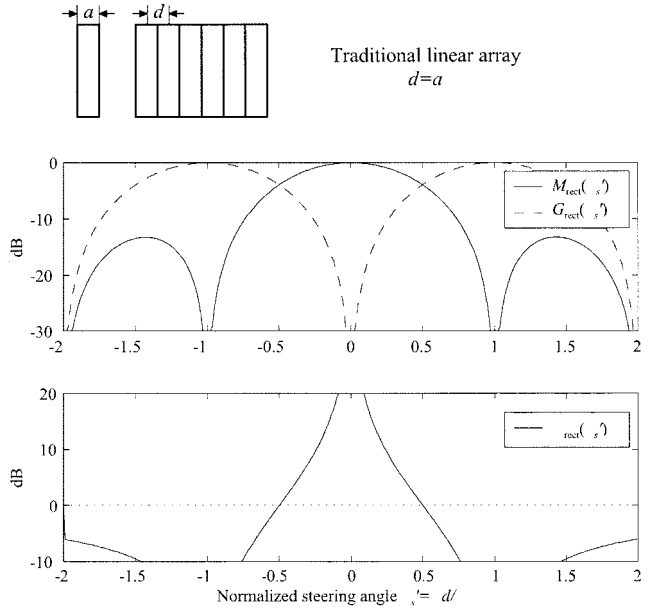


FIG. 4. A rectangular element dense linear array, where $d=a$. The upper plot shows the magnitude of the main lobe M and grating lobe G as a function of normalized steering angle β_s', and the bottom plot shows the ratio ζ of main lobe to grating lobe amplitude. Note that, while the magnitude of the main lobe does not diminish substantially for small angles, the main lobe to grating lobe ratio vanishes quickly for even small steering angles.

with the overall extent of the array D. As the characteristics of interest are due only to elements themselves, an infinitely large array is assumed, so that the lobes are impulses.

For a dense rectangular (traditional linear)array, where $d=a$, the main lobe magnitude M and grating lobe magnitude G as a function of steering angle are

$$M_{\mathrm{rect}}(\beta_s)=\mathrm{sinc}\,(\pi\beta_s a/\lambda), \tag{15}$$

$$G_{\mathrm{rect}}(\beta_s)=\mathrm{sinc}\,(\pi(\beta_s a/\lambda\pm 1)), \tag{16}$$

and the ratio ζ of the main lobe to grating lobe is

$$\zeta_{\mathrm{rect}}(\beta_s)=\frac{\mathrm{sinc}\,(\pi\beta_s a/\lambda)}{\mathrm{sinc}\,(\pi(\beta_s a/\lambda\pm 1))}. \tag{17}$$

In normalized angular coordinates $\beta_s'=\beta_s d/\lambda$,

$$M_{\mathrm{rect}}(\beta_s')=\mathrm{sinc}\,(\pi\beta_s'), \tag{18}$$

$$G_{\mathrm{rect}}(\beta_s')=\mathrm{sinc}\,(\pi(\beta_s'\pm 1)), \tag{19}$$

$$\zeta_{\mathrm{rect}}(\beta_s')=\frac{\mathrm{sinc}\,(\pi\beta_s')}{\mathrm{sinc}\,(\pi(\beta_s'\pm 1))}. \tag{20}$$

These results (in decibels) are plotted in Fig. 4. Note that, while the main lobe amplitude is strong for modest angles, the ratio ζ diminishes very quickly as the beam is steered. An improved phased array system would have a more uniform main lobe response and lobe ratio ζ for a desired steering sector. As these functions are wholly dependent on the element source function $s_0(x)$, this source function can be tailored to improve the lobe ratio ζ.

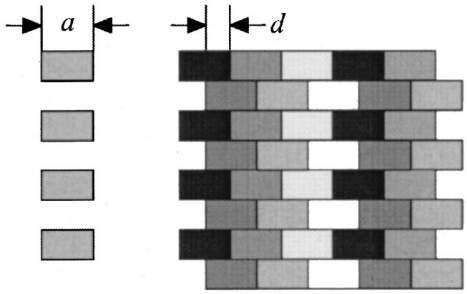


FIG. 5. Simple staggered element pattern. For clarity, each vertical element is shaded differently.

III. OVERSIZED ELEMENTS: STAGGERED ARRAYS

One way of improving the lobe ratio ζ is to narrow the directivity envelope $H_1(\beta)$ of the array. Narrowing this response will somewhat limit the maximum steering angle, but will further suppress grating lobes which occur at large angles. This necessitates increasing the element size compared to the interelement spacing. One of the ways to accomplish this in a planar array is to allow the interleaving or staggering of elements.

A. Staggered rectangular elements

A simple staggered array pattern is shown in Fig. 5. In this case, the individual element is a vertical set of rectangular apertures of width a driven as one unit. Because the elements are effectively overlapping, the element width is twice the interelement spacing d. Taking a polar cut along the x direction shows that the equivalent 1-D source function $s_0(x)$ is simply a rectangular window of width a.

The main and grating lobe magnitudes M and G, and their ratio ζ, are shown in Fig. 6. Clearly, this technique does

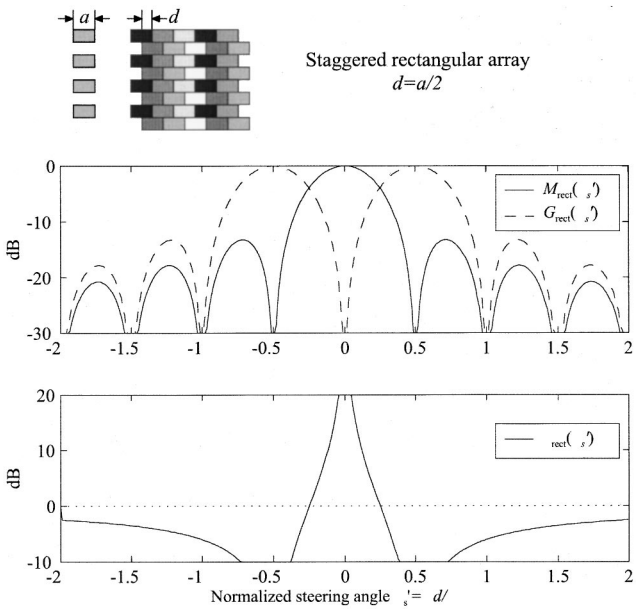


FIG. 6. The lobe magnitudes and main to grating lobe magnitude ζ for a rectangular staggered array, with interelement spacing equal to half the element width ($d=a/2$). Note that performance is diminished compared to the traditional linear array, both in main lobe magnitude and main to grating lobe ratio.

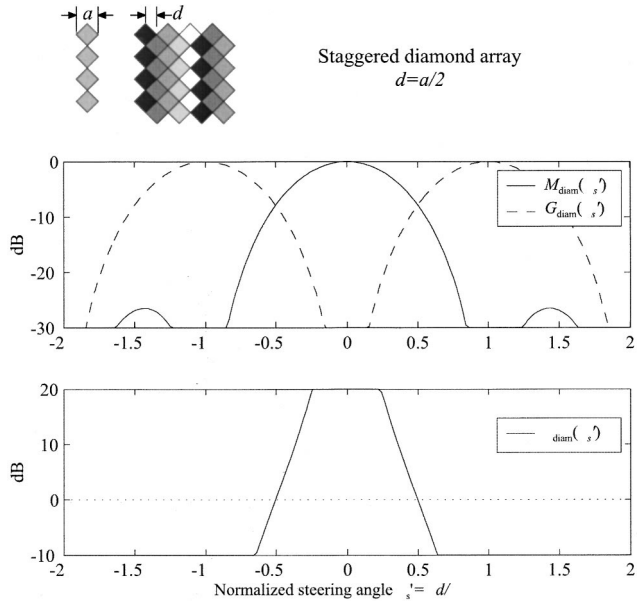


FIG. 7. A staggered linear array of diamond-shaped elements, with their more gradual taper, leads to a more uniform directivity envelope than the staggered rectangle array. There is also substantial improvement in main to grating lobe ratio ζ.

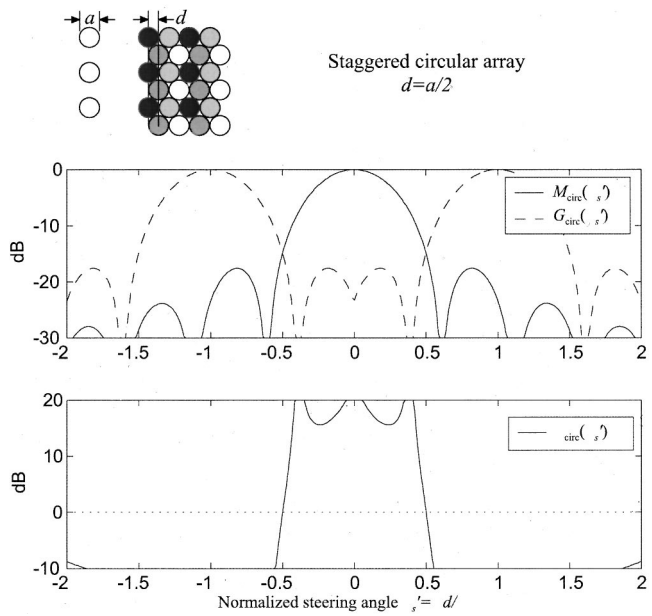


FIG. 8. The lobe magnitudes and main to grating lobe ratio ζ for a staggered circular array. While ζ is generally improved compared to that in the diamond array, the main lobe amplitude M is less uniform across steer angle.

narrow the array envelope, but does not improve the ratio ζ, due to the large ''side lobes'' now present in the directivity envelope. Furthermore, this technique significantly reduces the magnitude of the main lobe for even modest steering angles, which is of course undesirable.

B. Staggered diamond elements

The diamond pattern of array elements shown in Fig. 7 is worth considering, as it has a natural tightly packed configuration, and, being very similar to the rectangular array, is convenient to fabricate. Each element can be described as a triangular source function, where $\mathrm{tri}(a) = (-(2/a)x+1)\,\mathrm{sign}(x)$ for $-a/2 \leqslant x \leqslant a/2$, and zero otherwise. Its corresponding response is (with a scaling of the abscissa) that of the staggered rectangular response squared:

$$s_{\mathrm{diam}}(x) = \mathrm{tri}(a), \tag{21}$$

$$S_{\mathrm{diam}}(\beta) = [\mathrm{sinc}(\pi\beta_s a/2\lambda)]^2. \tag{22}$$

The more gentle taper in the element source function leads to a more uniform directivity envelope, leading to less attenuation of the main lobe across steering angle, as well as improvement in the lobe ratio ζ.

C. Staggered circular elements

The staggered circular element array, like the diamond element array, creates a more uniform directivity envelope due to its more gentle element taper. For this array, the source function and its response for a circular element of diameter a are simply

$$s_{\mathrm{circ}}(x) = \mathrm{circ}(a), \tag{23}$$

$$S_{\mathrm{circ}}(\beta) = \frac{J_1(\pi\beta a/\lambda)}{\pi\beta a/\lambda}, \tag{24}$$

where $\mathrm{circ}(a) = \sqrt{a^2/4 - x^2}$ for $-a/2 \leqslant x \leqslant a/2$ and zero otherwise, and J_1 is the Bessel function of the first kind.

In this case, the element width a is again twice the interelement spacing d. The large element size narrows the array response to an extent, but the more gradual taper of the circular aperture lowers side-lobe magnitude, and thus more effectively suppresses the grating lobe. This is clearly seen in the lower graph of Fig. 8.

For moderate steering angles ($|\beta_s'| < 0.4$), the lobe ratio ζ is superior to that of a rectangular or diamond shape, albeit at the expense of an attenuated main lobe. Here, the -3 dB main lobe angle is at $|\beta_s'| \approx 0.26$, compared to the corresponding $|\beta_s'| \approx 0.32$ for the diamond shape, or $|\beta_s'| \approx 0.45$ for a traditional rectangular element. In those applications for which the per channel cost is of greater concern than sensitivity, this can be a worthwhile tradeoff. However, for many applications where maximum sensitivity is important, this solution is inadequate. In the next section, an optimal element shape is developed not only to suppress grating lobes but also to maximize main lobe amplitude.

IV. OPTIMAL ELEMENTS

From the previous examples, it is shown that control of element shape, coupled with the use of staggered elements, improves the ratio of main lobe to grating lobe amplitude, at the expense of diminished main lobe amplitude. An optimal element shape would produce a directivity envelope corresponding to a rectangular window function, which is unity for $|\beta| < \beta_0$, and zero otherwise. For steering angles smaller than β_0, the main lobe would be unattenuated, but the grating lobe (at β_G) would be fully suppressed, so long as $|\beta_G| > \beta_0$. The corresponding element shape which produces such a function is, of course, the sinc function.

The difficulty in using the sinc function as an element shape lies in both its infinite spatial extent and the require-

ment of signal inversion to reproduce negative swings. Limiting the size of the element will diminish the sharpness of the directivity envelope, but through creative element design this effect can be minimized. Recall that because of the polar cut approximation in the one-dimensional array, the element source function of interest is actually the *vertically integrated* shape of the element, providing a great deal of flexibility in designing its actual shape. As far as the far-field approximation holds, the element can be offset or distributed as desired in the vertical dimension, allowing interleaving or other layout advantages.

The requirement of inverted areas of the elements may, depending on the specifics of the transducer, increase transducer complexity, and add challenges in signal routing. However, a signal inversion is usually easy to implement compared to generating a new delay line channel. In many cases, the electrical connections to the inverted elements can simply be reversed.

A. Altering the elements

The techniques presented in this study prescribe various element source functions, which can be implemented in a variety of ways. Most convenient for acoustic transducers are the following:

1. Element shading

It may be possible, depending on the application, to shade the response of each element, so that while the element itself maintains a traditional rectangular or other convenient shape, the source function $s_0(x)$ is no longer a simple rectangular aperture. For ultrasonic transducers, this could correspond to adding gradients in the thickness, density, or resonance mechanism along the vibrating element, or by deploying an acoustic filter to modify $s_0(x)$.

2. Element shaping

According to the polar cut approximation, an element driven uniformly will have an equivalent source function equal to the transverse integral across the element. Thus, an element can be shaped as needed to provide the desired response. This is most appropriate for a one-dimensional array.

Proper selection of shapes allows the use of staggered elements which are necessary for narrowing the array envelope. Care must be taken with effects along the vertical (nonazimuthal) direction, as the elements are no longer uniform along this axis; however, as long as this dimension is much larger than the element width, creating a very (vertically) narrow main lobe, these effects should be small.

3. Matrix approximation

In some transducer designs, such as diced-piezoelectric crystals or small microphones, it is inconvenient to shape the individual elements accurately. As an alternative, groups of small subelements may be gathered and cross-coupled to approximate a larger element having the desired source function. This may require a more complex interconnection scheme, but the reduced delay channel count may make this worthwhile.

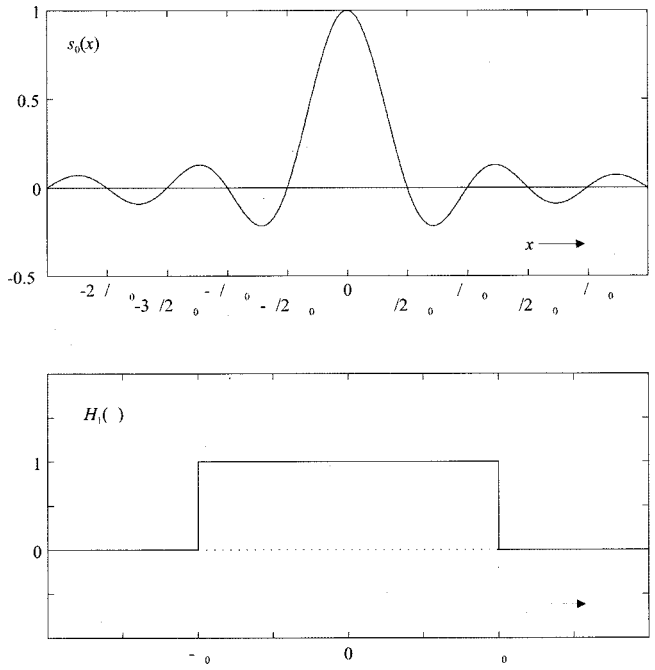


FIG. 9. An example of a sinc-type array element. The upper plot shows the element source function $s_{\text{sinc}}(x)=\text{sinc}\,(2\pi\beta_0 x/\lambda)$, which (if it were infinitely long) would produce the array envelope $H_{1,\text{sinc}}(\beta)=\text{rect}\,(2\beta_0)$, shown in the lower plot. By using the array envelope to limit maximum steering angle to β_0, grating lobes can be suppressed, allowing the relaxing of the interelement spacing restriction.

A practical implementation may use some combination of these methods. For a one-dimensional array, it is probably most straightforward to construct uniformly driven, specially shaped array elements.

B. Optimized element: Windowed sinc

The ideal element shape for producing an array envelope with cutoff angle β_0 is the sinc function, given by

$$s_{\text{sinc}}(x)=\text{sinc}\,(2\pi\beta_0 x/\lambda), \tag{25}$$

which has the corresponding response of

$$H_{1,\text{sinc}}(\beta)=\text{rect}\,(2\beta_0). \tag{26}$$

As shown in Fig. 9, the parameter β_0 corresponds to the dilation of the sinc function in the upper graph, whose zero crossings are at integer multiples of $\lambda/2\pi\beta_0$.

Recall that grating lobes are spaced at angles $\beta_G=\lambda/d$, so to avoid grating lobes completely, the grating lobes must be located at $|\beta_G|>\beta_0$, or

$$d<\frac{\lambda}{2\beta_0}. \tag{27}$$

Since $\beta_0<1$, this interelement spacing constraint is always less stringent than that of a traditional linear array. In fact, if the maximum steering angle is limited to ± 30 degrees ($\beta_0=0.5$), the interelement spacing, and therefore the required number of elements, can be decreased by half, without introducing grating lobes.

Because this type of element is not physically realizable, a windowed sinc function,

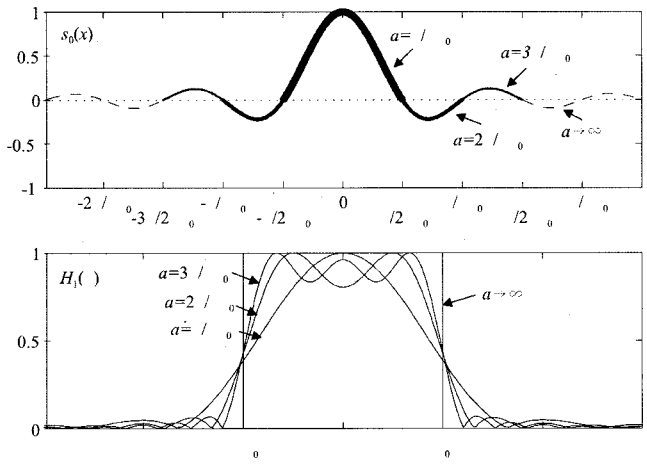


FIG. 10. Three proposed element source functions $s_0(x)$, each of which is a windowed sinc function, are shown in the upper graph. Below, their resulting array envelopes are shown. With a larger element size, the corresponding response approaches an ideal rectangular function.

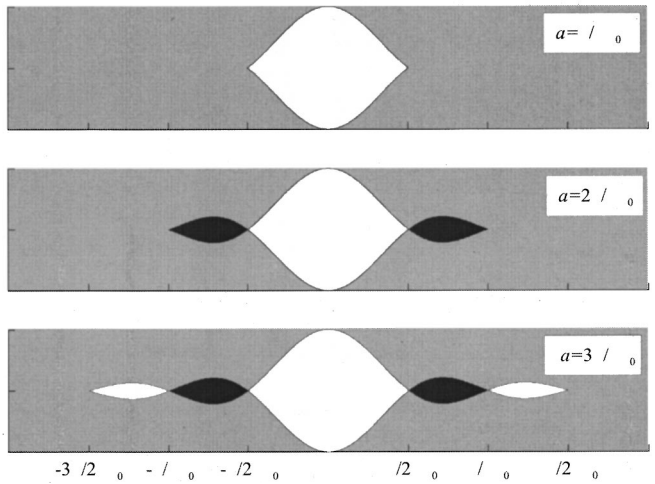


FIG. 11. Three sizes of the optimal sinc-shaped element are shown. The vertical dimension, having no substantial impact on beam steering, is in arbitrary units. The black areas, representing negative swings in the sinc function, are driven inverted with respect to the white areas.

$$s_{\mathrm{wsinc}}(x)=\mathrm{sinc}\,(2\pi\beta_0 x/\lambda)\cdot\mathrm{rect}\,a, \tag{28}$$

which is truncated to width a, may be used instead. This has the corresponding response:

$$S_{\mathrm{wsinc}}(\beta)=\mathrm{rect}\,(2\beta_0)*\mathrm{sinc}\,(\pi a\beta/\lambda). \tag{29}$$

The limitation of the length of the array element leads to a reduction in sharpness of the directivity envelope cutoff, as well as slight ripples in the main lobe magnitude for small steering angles. Of course, one could also use other window shapes (such as Chebyshev) to adjust these effects.

As it is logical to include an integral number of ''lobes'' (areas between zero crossings) in the element function, we can consider element lengths $a_0=\lambda/\pi\beta_0$, $a_1=2\lambda/\pi\beta_0$, and $a_2=3\lambda/\pi\beta_0$. These elements, with their corresponding response envelopes, are shown in Fig. 10.

C. Shaping elements

As discussed in an earlier section, the proposed element source function $s_0(x)$ can be obtained in several ways from a two-dimensional element, so long as

$$\int \hat{s}_0(x,y)\,dy=\mathrm{sinc}\,(2\pi\beta_0 x/\lambda)\cdot\mathrm{rect}\,(a).$$

The most straightforward method is to simply shape the elements in a two-dimensional plane as

$$\hat{s}_0(x,y)=\begin{cases}1 & \text{for } |y|\leqslant\mathrm{sinc}\,(2\pi\beta_0 x/\lambda)\cdot\mathrm{rect}\,(a)\\ 0 & \text{otherwise}\end{cases}. \tag{30}$$

This maintains symmetry in the vertical direction, and leads to a straightforward implementation. The element shapes for various a are shown in Fig. 11. While any desired length a can be used, it is logical to use element lengths which are bounded by zero crossings, as each of these areas specify discrete subelements.

Note that an arbitrary offset $y=v(x)$ can be added along the y direction without impacting the resulting integral because the response is integrated over the vertical direction. This allows a ''warping'' suitable for constructing staggered arrays. If desired, the shape can be reflected across the x axis, or subdivided into smaller segments, which can then be individually moved in the y direction. The impact of these transformations should be very small, even in along the elevation dimension, as long as the far-field approximation holds and the overall y dimension is much larger than the element size. Issues related to nonazimuthal effects will be explored in a later section.

D. Choice of element width

As has been discussed, the elimination of grating lobes requires an interelement spacing $d<\lambda/2\beta_0$. Note that the zero crossings of the element's sinc function are spaced in a manner proportional to λ/β_0. Therefore, the maximum allowable interelement spacing relative to the dilation of the source (sinc) function will be constant. In other words, any changes in the ratio λ/β_0 will simply rescale the x axis of the entire array, including the element itself and the interelement spacing, and limiting steerable angle.

In order to minimize the required number of elements by maximizing interelement spacing, the spacing $d=\lambda/2\beta_0$ can be used, which occurs slightly after the third zero crossing in the sinc function. This indicates that a convenient element width would be $a=3\lambda/\pi\beta_0$. In order to panel elements cleanly, d can be reduced slightly to $d=3\lambda/2\pi\beta_0$. The final array is shown in Fig. 12.

V. PERFORMANCE

The array proposed in Fig. 12 has an interelement spacing of

$$d=\frac{3\lambda}{2\pi\beta_0}, \tag{31}$$

which corresponds to a limiting steering angle of

$$\beta_0=\frac{3\lambda}{2\pi d}\approx\frac{\lambda}{2d}. \tag{32}$$

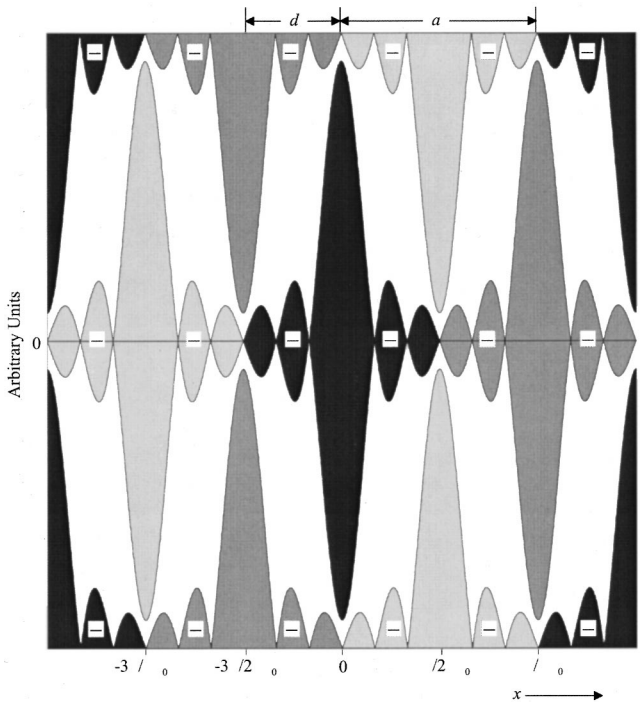


FIG. 12. Proposed array pattern, consisting of elements shaped as sinc functions. Each element is shaded differently to show their separation. Those regions whose excitation polarity is to be inverted are marked with a "–" sign. The horizontal axis is scaled relative to the ratio λ/β_0, and the vertical axis is of arbitrary units.

The element width is taken as $a=d/2$, as in the earlier interleaved designs. The resulting main lobe amplitude M_{sinc} and main lobe to grating lobe ratio G_{sinc} are shown Fig. 13. Clearly the array optimization has successfully suppressed

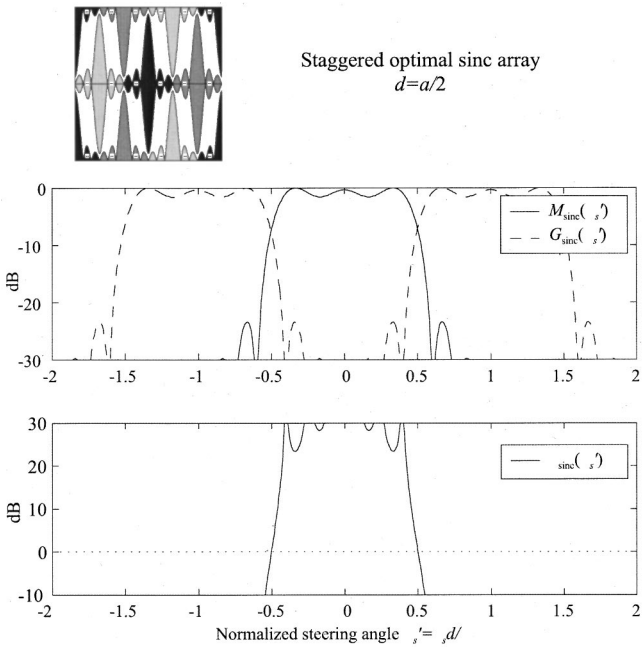


FIG. 13. Magnitude of main lobe, and main to grating lobe ratio, for the optimal array shown in Fig. 12. Steering angle is effectively limited to $|\beta_s'|<0.5$, over which the main lobe amplitude is nearly constant, and the grating lobe is effectively suppressed.

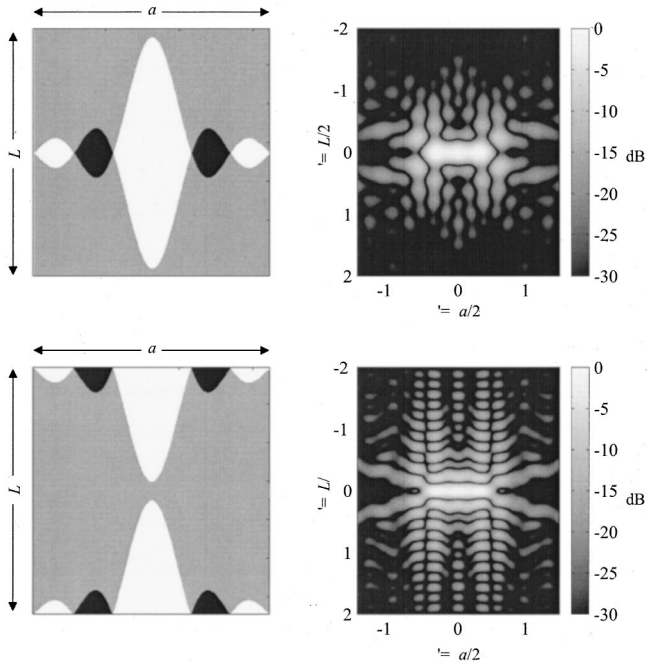


FIG. 14. The two styles of sinc elements are shown, with their respective responses, as a function of normalized angle. The black regions are driven inverted. For consistency with earlier figures, they are plotted against the normalized angles $\beta'=\beta a/2\lambda$ and $\gamma'=\gamma L/2\lambda$, where $a=2d$.

the grating lobe for limited steering angles, while maintaining a near-constant main lobe amplitude.

As with most phased array designs, this type of design must only meet the criteria for the smallest wavelength it is designed to handle, that is, for $\lambda \geqslant 2d\beta_0$. Any increase in λ simply increases the steering cutoff angle.

VI. TWO-DIMENSIONAL BEAM CHARACTERISTICS

From the previous section, the directivity envelope in the azimuthal plane approximates a rectangular function limited by steer angle β_0. Depending on the particular selection of tiling geometry, the vertical pattern of the envelope can vary significantly. For this reason, it is important that the performance, in particular the presence of spurious lobes, is assessed in the vertical plane.[13]

The analysis for the two-dimensional array is much like that for the one-dimensional array, but using a two-dimensional Fourier transform. The vertical size of the element can be defined as L, and elevation angle $\gamma=\sin\phi$.

Recall that, because of the polar cut approximation, the vertical dimension can be scaled, offset, or moved as desired, as long as the far-field approximation holds and the elements do not overlap. This allows an unlimited number of element tiling methods, although the most important has compact size, such as that shown in Fig. 12.

The linear array proposed in the previous section contains two types of elements, both with equivalent azimuthal responses. A two-dimensional analysis shows the expected azimuthal cutoff at $|\beta_0'|=0.5$ in Fig. 14, but the vertical characteristics may give rise to spurious vertical lobes if the main lobe is not sufficiently narrow. Because the vertical dimension L of the element makes equal contributions to the main

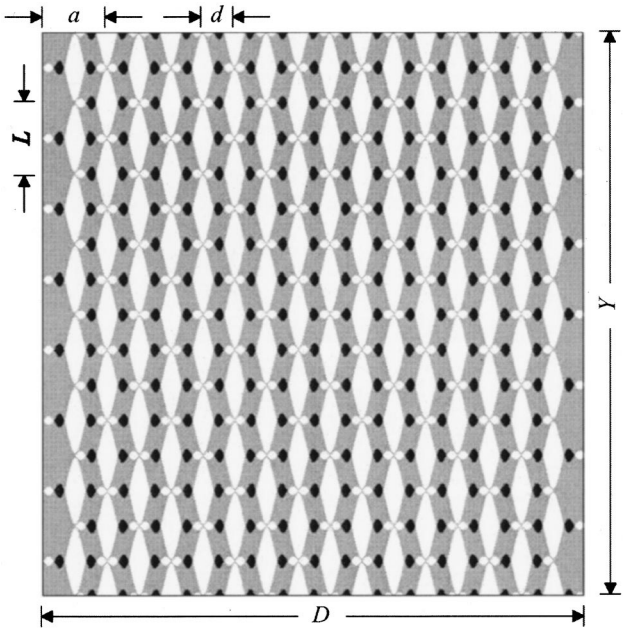


FIG. 15. The proposed linear array, using staggered sinc elements. Because the vertical dimension *Y* is much larger than the element height *L*, the main lobe will be sufficiently narrow to mitigate the presence of vertical side lobes. The interelement spacing, as before, is one-half the element size, in both the horizontal and vertical dimensions.

lobe height and the vertical envelope, simply scaling this parameter will not affect the presence of vertical lobes. Instead, if a two-dimensional staggered array of sinc elements, such as the one shown in Fig. 15, is used, no vertical sidelobes will develop. The vertical dimension *Y* of the array is much larger than the element, so the lobe height will be much narrower than the directivity envelope. Since the array is steered only in the azimuthal direction, grating lobes will never appear in the vertical direction.

VII. CONCLUSIONS

A unique and potentially useful method of phased array design has been presented, which puts particular emphasis on the shape of the individual elements in order to naturally suppress grating lobes. Much like a shaped or shaded array mitigates sidelobes, careful design of the constituent element shapes can naturally mitigate grating lobes, without detrimental effects to the main lobe.

By using properly shaped elements in staggered arrays, the minimum interelement spacing requirement can be smoothly exchanged for maximum steering angle. In many instances, this may lead to a substantial reduction in delay-line channels for a phased array system, as is particularly valuable in continuous-wave systems.

One limitation may be the cost and complexity of physically fabricating the proper transducer elements, or in inverting some subelements. While prohibitive for some applications, particularly in very-high-frequency ultrasound, there are many applications for which this may be a suitable and useful technique.

ACKNOWLEDGMENTS

This paper is a result of independent graduate term project. FJP would like to thank Professor Barry Vercoe and the Music, Mind & Machine Group for their helpful comments and general support. SCW would like to acknowledge the National Science Foundation for their encouragement (CMS-9903535).

[1] B. D. Steinberg, *Principles of Aperture and Array System Design* (Wiley, New York, 1976).
[2] S. C. Wooh and Y. Shi, "Optimum beam steering of linear phased arrays," Wave Motion **29**, 245–265 (1999).
[3] D. H. Turnbull and F. S. Foster, "Beam steering with pulsed two-dimensional transducer arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**(4), 320–333 (1991).
[4] B. D. Steinberg, "Comparison between the peak sidelobe of the random array and algorithmically designed aperiodic arrays," IEEE Trans. Antennas Propag. **AP-21**, 366–369 (1973).
[5] S. S. Brunke and G. R. Lockwood, "Broad-bandwidth radiation patterns of sparse two-dimensional vernier arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**(5), 1101–1109 (1997).
[6] S. W. Smith, H. G. Pavy, and O. T. von Ramm, "High-speed ultrasound volumetric imaging system—Part I: Transducer design and beam steering," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**(2), 100–108 (1991).
[7] G. R. Lockwood, P. C. Li, M. O'Donnell, and F. S. Foster, "Optimizing the radiation pattern of sparse periodic linear arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**(1), 15–19 (1996).
[8] G. R. Lockwood and F. S. Foster, "Optimizing the radiation pattern of periodic two-dimensional arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**(1), 15–19 (1996).
[9] O. T. von Ramm and S. Smith, "Beam steering with linear arrays," IEEE Trans. Biomed. Eng. **BME-30**(8), 438–452 (1983).
[10] F. J. Pompei, "The use of airborne ultrasound for generating audible sound beams," J. Audio Eng. Soc. **47**(9), 726–731 (1999).
[11] S. C. Wooh and Y. Shi, "Influence of phased array element size on beam steering behavior," Ultrasonics **36**, 737–749 (1998).
[12] S. P. Skobelev, "Methods of constructing optimum phased-array antennas for limited field of view," IEEE Antennas Propag. Mag. **40**(2), 39–50 (1998).
[13] S. C. Wooh and Y. Shi, "Three-dimensional beam directivity of phase-steered ultrasound," J. Acoust. Soc. Am. **105**, 3275–3282 (1999).

Full-wave modeling of therapeutic ultrasound: Nonlinear ultrasound propagation in ideal fluids

Siegfried Ginter,[a] Marko Liebler, Eckard Steiger,[b] Thomas Dreyer, and Rainer E. Riedlinger

Institut für Höchstfrequenztechnik und Elektronik/Akustik, Universität Karlsruhe, Kaiserstrasse 12, 76128 Karlsruhe, Germany[c]

(Received 1 May 2001; accepted for publication 3 January 2002)

The number of applications of high-intense, focused ultrasound for therapeutic purposes is growing. Besides established applications like lithotripsy, new applications like ultrasound in orthopedics or for the treatment of tumors arise. Therefore, new devices have to be developed which provide pressure waveforms and distributions in the focal zone specifically for the application. In this paper, a nonlinear full-wave simulation model is presented which predicts the therapeutically important characteristics of the generated ultrasound field for a given transducer and initial pressure signal. A nonlinear acoustic approximation in conservation form of the original hydrodynamic equations for ideal fluids rather than a wave equation provides the base for the nonlinear model. The equations are implemented with an explicit high-order finite-difference time-domain algorithm. The necessary coefficients are derived according to the dispersion relation preserving method. Simulation results are presented for two different therapeutic transducers: a self-focusing piezoelectric and one with reflector focusing. The computational results are validated by comparison with analytical solutions and measurements. An agreement of about 10% is observed between the simulation and experimental results. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1468876]

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

I. INTRODUCTION

The different therapeutic effects of intense ultrasound (US) like stone destruction, cavitation, and tissue heating by US absorption, together with the ability of ultrasound to reach deeply situated tissue in a noninvasive way, are the foundation of several medical applications. Today, the clinical application of therapeutic ultrasound is established in lithotripsy.[1] It is currently evaluated for orthopedic therapies[2] and for the treatment of prostatic diseases.[3] Recently, a number of new applications such as the noninvasive thermal ablation therapy of tumors[4] or the occlusion of blood vessels[5] were proposed. All these therapeutic applications use high-intense, focused ultrasound (HIFU).

As shown in Fig. 1, the effects and efficiency of these therapies are directly related to the field parameters in the focal area inside the patient's body, such as the maximum and minimum amplitudes, their spatial distribution, and the waveforms. Only the knowledge of these field parameters improves the understanding of the interaction between the ultrasound wave and the biological tissue, and is therefore essential for controlling and improving the therapy. Furthermore, the ultrasonic field in the focal area is determined by the design of the therapy system (focusing gain, signal intensity, aperture, excitation waveform, frequency) and the propagation properties, like nonlinearity and attenuation, of the overlying medium layers (water, tissue). Measurements in therapeutically used transducers clearly show the existence of nonlinear effects for ultrasound propagation, which lead to a steepening of the wave profile. For high signal intensities and strong focusing sources, the nonlinear steepening results in the formation of a shock wave. The first step towards an understanding of the relation between the parameters at the US source and the achieved therapeutical effects consists of the investigation of the physically describable US propagation (gray box in Fig. 1). Therefore, an accurate simulation tool which takes nonlinear effects into account is inevitable.

Linear models are not able to accurately predict the important field parameters, and are therefore inadequate to simulate therapeutic ultrasound. Most nonlinear-acoustic models are based on a second-order approximation of the complete nonlinear equations of fluid motion.[6,7] As a model equation for the propagation of finite-amplitude sound for therapeutic applications, Hallaj[8] uses the Westervelt equation, which accounts for nonlinear propagation of locally plane waves in a thermoviscous fluid.[7] In solving the wave equation in the limit of a lossless fluid with a standard finite-difference algorithm in the time domain (FDTD), the correct reproduction of weak shocks, which occur in therapeutic ultrasound, are not guaranteed. Another method to model therapeutic ultrasound is proposed by Wojcik.[9] It is based on a time-domain finite-element method that solves the partial differential equations for the US propagation in a linear approximation. Only cumulative nonlinear steepening effects are included by a nonlinear material relationship between pressure and density. The propagation of strongly nonlinear waves in an ideal gas is investigated by Yano and Inoue[10] for a harmonic oscillating piston source. They solve the Euler equations with an explicit high-resolution upwind finite-

[a]Electronic mail: Siegfried.Ginter@etec.uni-karlsruhe.de
[b]Now at: Robert Bosch GmbH, P.O. Box 30 02 40, D-70442 Stuttgart, Germany.
[c]URL http://www-ihe.etec.uni-karlsruhe.de

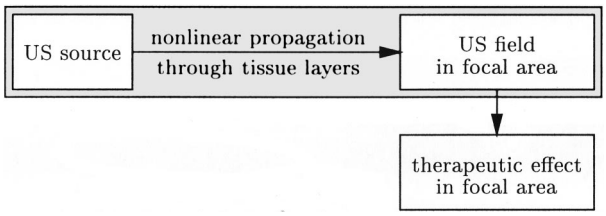


FIG. 1. The causal relationship between the therapy system and the therapeutic effects.

difference scheme. The most common approach today is to solve the KZK equation numerically, as used by Hamilton *et al.*,[11,12] Bernsten and Ystad,[13] and Tjøtta and Tjøtta.[14] Introduced by Khokhlov, Zabolotskaya, and Kuznetsov,[6] KZK describes a parabolic approximation of the nonlinear acoustic wave equation and consistently accounts for the combined effects of nonlinearity, diffraction, and absorption, but only for directional sound beams and US propagation in homogeneous media.[7,15] The approximation assumes that the transverse field variations are slow compared to longitudinal variations along the acoustic axis. For strong focusing transducers this assumption fails, as it does for inhomogeneous media or reflector focusing therapy systems. Rudenko and Sukhorukova[16] presented an extended model for propagation of directional sound beams in inhomogeneous media. If one nevertheless wants to simulate reflector focusing[17] or strong focusing systems[13] with the KZK equation, supplementary propagation models are needed which provide the initial conditions for the subsequent KZK propagation modeling. Finally, a model of nonlinear diffractive field propagation, which is not restricted to the parabolic wave approximation, was developed by Christopher and Parker.[18] Using an operator-splitting method, this model takes into account nonlinear effects in the temporal frequency domain and diffraction and attenuation in the spatial transform domain. With an updated version of this model, Christopher considered the field propagation in an electrohydraulic shock wave lithotripter.[19] To avoid numerical errors associated with the transformations between the time and frequency domains, Tavakkoli *et al.*[20] presented a pure time-domain model, based on the approach used by Christopher.

In this paper, a numerical model originating from the equations of hydrodynamics is used. It includes the propagation of nonlinear ultrasound waves and weak shocks in ideal fluids for axisymmetric focusing transducers and can be extended to inhomogeneous media. No restrictions on beam direction or the focusing systems are imposed; therefore, it is called a full-wave model which inherently includes the effects of diffraction, refraction, and reflection. The model is finally solved with an explicit FDTD scheme. The simulation results are used to characterize the ultrasound field of different therapy systems in pure degassed water, which behaves nearly like an ideal fluid. In a subsequent paper an extension to this new model treats the propagation through attenuating tissue layers.

After a description of the mathematical model in Sec. II, the numerical implementation is shown in Sec. III. In Sec. IV the simulation results are validated by comparison with an analytical solution and with measurements done in two different US transducers. Finally, some conclusions follow in Sec. V.

II. MATHEMATICAL MODEL

The mathematical description of the full-wave US propagation model that includes the nonlinear steepening and the formation and propagation of weak shocks, starts from the general equations of hydrodynamics for ideal fluids.[21] In an ideal fluid the effects of viscosity and thermal conduction are neglected and therefore no dissipation of energy occurs during US propagation.[21] The hydrodynamic equations (1) describe the temporal change of mass in a fixed volume V due to the mass flux across a closed boundary surface S in Eq. (1a) and the change of momentum for a moving volume V^* in Eq. (1b) by neglecting the influence of gravity

$$\frac{\partial}{\partial t}\int\int\int_V \rho_g \,\mathrm{d}V = -\oint_S \rho_g \mathbf{v}_g \mathbf{n}\,\mathrm{d}S, \tag{1a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\int\int_{V^*} \rho_g \mathbf{v}_g \,\mathrm{d}V = -\oint_{S^*} p_g \mathbf{n}\,\mathrm{d}S. \tag{1b}$$

Here, t is the time, ρ_g the density of the fluid, $\mathbf{v}_g$ the velocity of a fluid particle, and p_g the pressure normal to the surface S ($\mathbf{n}$ represents the outward-directed normal vector to the surface). Assuming that the fluxes are continuously differentiable, the application of the Gauss integral identity in the limit of $V \to 0$ leads to a differential formulation. Applying the identity Eq. (2) to Eq. (1b)[22]

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int\int\int_{V^*} \rho_g \mathbf{v}_g \,\mathrm{d}V &= \int\int\int_{V^*} \rho_g \frac{\mathrm{d}\mathbf{v}_g}{\mathrm{d}t}\,\mathrm{d}V \\ &= \int\int\int_{V^*} \rho_g \left[\frac{\partial \mathbf{v}_g}{\partial t} + (\mathbf{v}_g \nabla)\mathbf{v}_g\right]\mathrm{d}V,\end{aligned} \tag{2}$$

Eq. (1) then becomes

$$\frac{\partial \rho_g}{\partial t} + \nabla(\rho_g \mathbf{v}_g) = 0, \tag{3a}$$

$$\rho_g\left[\frac{\partial \mathbf{v}_g}{\partial t} + (\mathbf{v}_g \nabla)\mathbf{v}_g\right] = -\nabla p_g. \tag{3b}$$

When treating acoustic problems, it is convenient to split the quantities p_g, ρ_g, and $\mathbf{v}_g$ into a constant part (index 0) and a varying part. With $\mathbf{v}_0 = 0$, the perturbation approach in Eq. (4) introduces the acoustic variables: the acoustic density ρ, the acoustic pressure p, and the acoustic velocity $\mathbf{v}$,

$$\rho_g = \rho_0 + \rho, \quad p_g = p_0 + p, \quad \mathbf{v}_g = \mathbf{v}. \tag{4}$$

Inserting Eq. (4) into Eq. (3), the nonlinear acoustic equations (5) follow

$$\frac{\partial \rho}{\partial t} + \nabla(\rho_0 + \rho)\mathbf{v} = 0, \tag{5a}$$

$$(\rho_0+\rho)\left[\frac{\partial \mathbf{v}}{\partial t}+(\mathbf{v}\nabla)\mathbf{v}\right]=-\nabla p. \tag{5b}$$

Neglecting the higher-order terms in the acoustic variables, the linear acoustic equations (6) result

$$\frac{\partial \rho}{\partial t}=-\nabla(\rho_0 \mathbf{v}), \tag{6a}$$

$$\frac{\partial \mathbf{v}}{\partial t}=-\frac{1}{\rho_0}\nabla p. \tag{6b}$$

Together with a linear relationship between pressure and density, Eq. (6) is directly related to the well-known linear acoustic wave equation. Modeling the propagation of high-intense US, the linear equations in Eq. (6) are not sufficient. In order to include steepening of the US wave, the nonlinear terms in Eq. (5) and the nonlinearity of the medium must be included. These terms describe an amplitude-dependent phase velocity of the wave. As described by Šutilov,[23] points in the wave profile with a large amplitude propagate faster than points with a low amplitude. In the spectral domain this wave distortion can be described by a transfer of energy from the fundamental frequency component to higher harmonics. The pressure gradient at the front of the wave becomes more and more steep, leading to a high acoustic velocity gradient. Because all real fluids show at least a low viscosity, this high velocity gradient leads to strong local frictional forces. Finally, a stable state, called shock, exists between the counteracting processes of nonlinear steepening and dissipation. Measurements by Eisenmenger[24] determine the typical thickness of the shock front in water to be on the order of 10^{-6} m.

The ideal nonlinear acoustic equations (5) do not include the shock formation and propagation correctly. In absence of the dissipative processes the gradient at the front becomes infinite and the description by the partial differential equation (PDE) fails. To model especially the amplitude and the propagation speed of shocks, the ideal fluid model has to be extended.

One approach explicitly includes the dissipative effects at the shock front. By this, the gradient in the front remains finite and the PDE description is therefore valid. A discrete algorithm, like the FDTD, for a direct modeling of the dissipative effects inside the small-sized shock front seems questionable because of the exhaustive computational costs involved.

In the following, the basic mathematical background is explained for our new model that predicts accurately nonlinear steepening effects and the formation and propagation of weak shocks. If the exact curvature of the shock front is of minor interest compared to the correct amplitude and propagation speed of the shock, another path can be taken: shocks are treated as discontinuities having an infinite gradient. Mathematically they could be regarded as weak solutions of the PDE. To include such weak solutions in a correct manner, one has to use the original integral equations (1). Applied on a discontinuity, these integral equations lead to jump conditions for the field quantities. Together with an energy equation, these are the so called Rankine–Hugoniot relations, which have to be fulfilled at the discontinuity. They describe the quantities that are continuous across a shock, namely the mass flux, the momentum flux, and the flux of energy. From the second law of thermodynamics, it follows that the entropy increases when passing through the discontinuity,[22] which is equivalent to dissipation in physical terms. This is often called the entropy condition. The amount of this dissipation is nearly independent of the kind and strength of the physical dissipation processes in the shock front and depends only on the conservation law, used on both sides of the discontinuity.[22] Hence, independently of the physical kind and strength of the dissipation processes in the shock front, the fulfillment of the jump conditions provide the correct amplitude and speed of the shock.[21]

The extended model for nonlinear propagation in ideal fluids consists of the above-mentioned PDE, which holds in regions of continuous solutions. At discontinuities the model has to be supplemented by the jump conditions. To ensure that the numerical solution converges to a physical one, a numerical viscosity term is appended to fulfill the entropy condition. More details on the numerical implementation of this extended model are given in Sec. III.

To close the system of equations, an equation of state $p=p(\rho,s)$ has to be added, with s being the entropy variation due to dissipative effects. The entropy variation term in the state equation can be estimated. Assuming Mach numbers $M \ll 1$ ($M=|v|_{\max}/c_0$) for all therapeutic US fields, the entropy variations due to thermal conductivity of water and due to the dissipative effects inside the shock front can be neglected. According to the theory of weak shocks, the dissipation is proportional to the third power of the pressure difference at the shock front.[7,21,22,25] In an approximation valid up to the second order, the adiabatic state equation for homogeneous media can be written as

$$p=c_0^2\rho+\frac{c_0^2}{\rho_0}\frac{B}{2A}\rho^2, \tag{7}$$

where c_0 represents the sound speed and B/A the acoustic nonlinearity parameter of the medium. In a linear approximation of Eq. (7), one obtains

$$p=c_0^2\rho. \tag{8}$$

From the considerations above it follows, that the extended ideal fluid model, consisting of Eqs. (5) and (7) in a second-order approximation completed by the jump conditions and the entropy condition, is necessary and sufficient for the description of therapeutic ultrasound propagation, including weak shocks. No complementary energy equation is needed.

III. NUMERICAL IMPLEMENTATION

For a numerical implementation, a two-dimensional explicit FDTD algorithm is chosen to solve Eqs. (5) and (7) in cylindrical coordinates, assuming axisymmetry. The explicit FDTD algorithm is a simple, efficient, and flexible method to calculate US wave propagation including nonlinear steepening effects in complex geometries.

The algorithm has to take into account the different possible types of solutions of the problem. Over wide parts of

the computational domain the solutions are expected to behave smoothly. Caused by the nonlinear steepening process, weak shocks are generated mainly in the focal region, which have to be taken into consideration, too. In order to include this huge time scale ''efficiently,'' a two-stage strategy is applied.[25] To include weak shocks in a correct manner, a numerical method in conservation form is derived in Sec. III A,[26,27] which then is implemented by a dispersion relation preserving (DRP) high-order finite-difference scheme[28] for an optimal treatment of broadband smooth solutions, presented in Sec. III B.

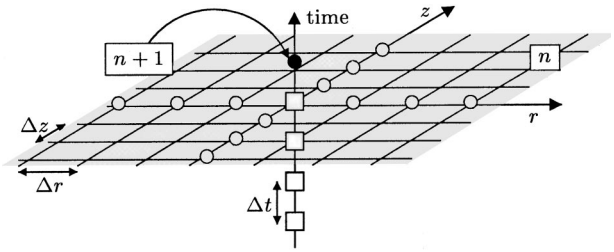


FIG. 2. The applied explicit, high-order FDTD scheme. For spatial differentiation a symmetric, seven-point stencil (open circles) is used in r- and z-directions. Integration in time (open squares) is over four time levels for the new field value (black circle).

A. Numerical treatment of weak shocks

In order to guarantee a correct representation of weak-shock solutions with a numerical method based on partial differential equations, the conditions of the Lax–Wendroff theorem[26] must be fulfilled: A numerical method in conservation form guarantees that the solution converges to the correct discontinuous weak solution, when reducing the step size towards zero, provided that the method is consistent. One way to derive a numerical method in conservation form is to use standard finite-difference discretization and start with the conservative form of the PDE, not using Eq. (5) directly.[25,26] With

$$\frac{1}{(\rho_0+\rho)}=\frac{1}{\rho_0}-\frac{\rho}{\rho_0(\rho_0+\rho)}, \tag{9}$$

and the divergence operator in two-dimensional cylindrical coordinates (r,z), one obtains from Eq. (5) the following conservation law formulation:

$$\frac{\partial \mathbf{w}}{\partial t}+\frac{\partial \mathbf{F}^r}{\partial r}+\frac{\partial \mathbf{F}^z}{\partial z}=\mathbf{Q},$$

with

$$\mathbf{w}=\begin{pmatrix}\rho(r,z,t)\\ v_r(r,z,t)\\ v_z(r,z,t)\end{pmatrix},$$

$$\mathbf{F}^r=\begin{pmatrix}(\rho_0+\rho)v_r\\ p/\rho_0\\ 0\end{pmatrix},\quad \mathbf{F}^z=\begin{pmatrix}(\rho_0+\rho)v_z\\ 0\\ p/\rho_0\end{pmatrix}, \tag{10}$$

$$\mathbf{Q}=\begin{pmatrix}-\dfrac{(\rho_0+\rho)v_r}{r}\\ \dfrac{\rho}{(\rho_0+\rho)\rho_0}\dfrac{\partial p}{\partial r}-\left(v_r\dfrac{\partial}{\partial r}+v_z\dfrac{\partial}{\partial z}\right)v_r\\ \dfrac{\rho}{(\rho_0+\rho)\rho_0}\dfrac{\partial p}{\partial z}-\left(v_r\dfrac{\partial}{\partial r}+v_z\dfrac{\partial}{\partial z}\right)v_z\end{pmatrix}.$$

The temporal changes in the conservative quantities $\mathbf{w}$, are caused by the sum of the spatial changes of the fluxes $\mathbf{F}^r$ and $\mathbf{F}^z$. The source vector $\mathbf{Q}$ contains the inhomogeneous part of the PDEs, i.e., some nonlinear and geometrical terms, which cannot be included into the flux terms. The first equation in the set of equations (10) represents a conservation equation for the flux of mass and therefore leads to the corresponding Rankine–Hugoniot relation. The second and third equation in the numerical conservation form (10) approximate the conservation of momentum flux in a linear order. Therefore, the set of equations (10) provides a first-order approximation of the exact shock propagation speed, which is sufficient for most practical applications.

The fulfillment of the Lax–Wendroff theorem ensures a correct propagation speed of a discontinuous weak solution independently of the discretization. To ensure the correct amplitudes, the discretization has to be chosen so that discontinuities are well resolved. However, this theorem does not guarantee that the weak solution is a physical one. A numerical, artificial viscosity term incorporated into Eq. (10) ensures that the entropy behaves physically and increases over the discontinuity,[26]

$$\frac{\partial \mathbf{w}}{\partial t}+\frac{\partial \mathbf{F}^r}{\partial r}+\frac{\partial \mathbf{F}^z}{\partial z}=\mathbf{Q}-\mu\frac{\partial^2 \mathbf{w}}{\partial r^2}-\mu\frac{\partial^2 \mathbf{w}}{\partial z^2}, \tag{11}$$

with μ as the artificial viscosity. Now, the basic formulation of the PDE is prepared in such a manner that an FDTD implementation includes not only smooth but also discontinuous solutions.

B. Numerical treatment of acoustic solutions—the applied FDTD scheme

For an optimal treatment of nonlinear acoustic solutions, an explicit high-order FDTD algorithm is chosen to implement Eq. (11) numerically. The continuous differential operators are approximated by discrete difference operators with equidistant spatial discretization $\Delta r=\Delta z$. The solution $\mathbf{w}_{i,j}^n$ is calculated for each time step $t_n=n\Delta t$ in each discrete grid point $(r_i=i\Delta r, z_j=j\Delta z)$. An explicit algorithm yields the solution $\mathbf{w}_{i,j}^{n+1}$ of the next time step by integrating over calculated quantities from preceeding time levels. Figure 2 illustrates the applied FDTD scheme. To get $\mathbf{w}_{i,j}^{n+1}$, an Adams–Bashforth method is used to integrate the temporal changes of $\mathbf{w}_{i,j}(t)$. These temporal changes of $\mathbf{w}_{i,j}(t)$ are approximated by a sum of the spatial derivatives of the flux quantities $\mathbf{F}^r$ and $\mathbf{F}^z$, the source terms $\mathbf{Q}$, and the terms of the artificial viscosity. Here, this sum will be denoted as $\mathbf{H}_{i,j}^n$. Finally, the following FDTD implementation of Eq. (11) is obtained:

$$\mathbf{w}_{i,j}^{n+1} \approx \mathbf{w}_{i,j}^{n} + \Delta t \sum_{r=0}^{r=3} b_r \left.\frac{\partial \mathbf{w}}{\partial t}\right|_{i,j}^{n-r}$$

$$\approx \mathbf{w}_{i,j}^{n} + \Delta t \sum_{r=0}^{r=3} b_r \mathbf{H}_{i,j}^{n-r}, \tag{12}$$

with

$$\mathbf{H}_{i,j}^{n} \approx -\frac{1}{\Delta r}\sum_{p=-3}^{3} a_p \mathbf{F}_{i+p,j}^{r,n} - \frac{1}{\Delta z}\sum_{p=-3}^{3} a_p \mathbf{F}_{i,j+p}^{z,n} + \mathbf{Q}_{i,j}^{n}$$

$$- \nu \sum_{s=-3}^{3} c_s \mathbf{w}_{i+s,j}^{n} - \nu \sum_{s=-3}^{3} c_s \mathbf{w}_{i,j+s}^{n}, \tag{13}$$

where b_r are the coefficients for the numerical integration, a_p are the coefficients for the numerical derivative of the fluxes $\mathbf{F}$, c_s are to approximate the numerical artificial viscosity terms, and finally $\nu = \mu/\Delta x^2$ is the constant parameter of the artificial viscosity. The approximation of the spatial derivatives is performed with a symmetric seven-point stencil, and for the temporal integration the last four time levels are used. The coefficients a_p and b_r are designed according the DRP (dispersion relation preserving) method.[28] The aim of this method is to preserve the dispersion relation of the underlying PDE in its numerical representation. To get DRP-optimized coefficients a_p and b_r, two different methods for coefficient design, namely the method of the Taylor series expansion and the method of coefficient design in the frequency domain, are combined. So, the DRP method combines the advantage of being consistent as $\Delta x \to 0$ from the Taylor method with the broadband approximation behavior typical for the frequency domain method. The coefficients b_r and a_p are listed in the Appendix. For a further description, see Refs. 28 and 25.

To study the numerical propagation behavior of the DRP method, a plane-wave expansion of the discrete differentiation and integration operators in Eq. (13) and Eq. (12) is considered. With plane waves of the form $W(x,t) = \hat{W} e^{j(kx-\omega t)}$, where $\hat{W}$ is the complex amplitude, k the wave number, ω the angular frequency, and x the propagation direction, an approximated numerical wave number $\overline{k}\Delta x$ relative to the spatial grid is obtained as a function of the normalized physical wave number $k\Delta x$

$$\overline{k}\Delta x = -j \sum_{p=-3}^{p=3} a_p e^{jpk\Delta x}, \tag{14}$$

and for the normalized approximated angular frequency referred to the temporal discretization, $\overline{\omega}\Delta t$, as a function of $\omega \Delta t$

$$\overline{\omega}\Delta t = \frac{j(e^{-j\omega\Delta t} - 1)}{\sum_{r=0}^{r=3} b_r e^{jr\omega\Delta t}}. \tag{15}$$

Using the values of the coefficients a_p and b_r mentioned in the Appendix, for a smaller maximum difference $|\overline{k}\Delta x - k_{\max}\Delta x|$ of 0.005, a maximal normalized physical wave number $k_{\max}\Delta x$ is extracted from Eq. (14). The value of $k_{\max}\Delta x = 1.17$ also means that in theory 5.4 grid points per wavelength are sufficient for a correct discrete spatial representation of the wave. Also, from Eq. (14) the maximal value of the approximated normalized wave number can be calculated to $\overline{k}_{\max}\Delta x = 1.65$, which indicates the maximal possible wave number $\overline{k}$ occurring in the numeric calculation. The evaluation of Eq. (15) shows four different complex solutions $\omega_{1,2,3,4}$ for a given numerical $\overline{\omega}$. To avoid instabilities, the imaginary parts of all possible $\omega\Delta t$ curves have to be less than or equal to zero (no unphysical amplification) and so $\overline{\omega}_{\min}\Delta t < 0.4$ must be fulfilled.

The artificial numerical viscosity $\mu = \nu \Delta x^2$ in the implementation of Eq. (11) quadratically decreases towards zero as $\Delta x \to 0$. So, the correct weak solution is guaranteed in the limit of a vanishing cell size. The coefficients c_s are determined in the frequency domain to approximate a spatial Gaussian low-pass filter. The low-frequency parts of the solution remain undamped. Only waves with $k_{\max}\Delta x > 1.17$, which do not propagate correctly across the numerical grid, are selectively damped. Here, $\nu = 0.1$ is used. To avoid effects of signal distortion due to numerical dispersion, a symmetric difference scheme is used. The filter coefficients c_s are listed in the Appendix.

Besides the consistency of the numerical method, the stability is important. Every explicit algorithm for hyperbolic equations has to fulfill the Courant–Friedrichs–Levy stability condition, which couples the maximal allowed time step to the spatial discretization.[26] The factor γ is called the Courant number. For the DRP algorithm in two dimensions, one gets the following relation:

$$c_0 = \gamma \frac{\Delta x}{\Delta t}, \quad \text{with} \quad \gamma \leqslant \frac{\overline{\omega}_{\min}\Delta t}{(M+\sqrt{2})\overline{k}_{\max}\Delta x} < 1. \tag{16}$$

Assuming Mach numbers M smaller than 0.2 for all therapeutic applications, γ is required to be smaller than 0.149.

C. Numerical treatment at the boundary

A numerical model for the calculation of US propagation in therapy systems has to consider various conditions at the boundary of the numerical grid. At the surface of the transducer, the US signal is emitted. Here, a source condition has to be implemented. The surrounding of the active area, as the mounting or the reflector, is most often made of metal like brass or aluminum. In an approximation, these boundaries can be regarded as ideally rigid: there is no normal velocity component $v_\eta = 0$ and no normal pressure gradient $\partial p/\partial \eta = 0$ at the boundary. This condition is also applied to model the boundary condition on the symmetry axis. If the boundary doesn't behave ideally rigid or soft, impedance boundary conditions are necessary. Finally, absorbing boundary conditions are needed to limit the computational domain, where no physical boundary exists.

For simplicity reasons, all considerations at the boundary are made in a local coordinate system η and ξ, where η stands for the normal and ξ for the tangential direction at the boundary. The transform of the local η,ξ system to the global r,z system is executed by a rotation matrix.

For the implementation of ideally rigid (or ideally soft) boundary conditions, the image method is applied. As shown in Fig. 3, the numerical grid is extended by three dummy

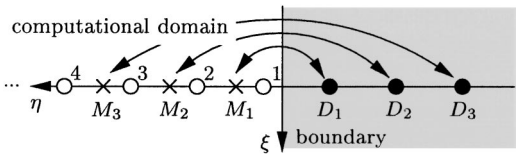


FIG. 3. Implementation of ideally rigid and ideally soft boundary conditions. The inner grid points (1,2,3,...) are expanded with three outer dummy points (D_1, D_2, D_3), values of which are determined by the corresponding image points (M_1, M_2, M_3).

points (index D). So, the full high-order symmetric FDTD scheme is applicable throughout the whole computational domain. The quality of the numerical method remains the same at the boundary as in the inner region. The values of the field variables (ρ, v_η, v_ξ) in the three dummy points can be calculated from the interpolated values of the corresponding image points (index M) by the following relation:

$$\begin{aligned} &\text{rigid:} \quad (\rho, v_\eta, v_\xi)_D = (\rho, -v_\eta, v_\xi)_M, \\ &\text{soft:} \quad (\rho, v_\eta, v_\xi)_D = (-\rho, v_\eta, v_\xi)_M. \end{aligned} \tag{17}$$

Reflector focusing transducers especially need accurate modeling of the curved rigid reflector boundary. There, the values in the image points are calculated by a two-dimensional local spline interpolation, as described in Refs. 25 and 29.

The implementation of the impedance boundary condition, the absorbing boundary condition, and the source condition is based on a characteristic form of Eq. (5). These conditions are expressed quite easily by separating incoming and outgoing waves.[25,30–33] To derive the boundary conditions in characteristic form, the consideration starts with the quasilinear form of Eq. (5) in local coordinates, Eq. (18)

$$\frac{\partial \mathbf{w}}{\partial t} + A \frac{\partial \mathbf{w}}{\partial \eta} + \mathbf{F}_{\text{trans}} = 0, \tag{18}$$

with the vector $\mathbf{w} = (\rho, v_\eta, v_\xi)$. The vector $\mathbf{F}_{\text{trans}}$ contains the transversal and geometrical terms of Eq. (5), which are not in $A \partial \mathbf{w}/\partial \eta$, with

$$A = \begin{bmatrix} v_\eta & (\rho_0 + \rho) & 0 \\ \dfrac{c_0^2}{(\rho_0 + \rho)} & v_\eta & 0 \\ 0 & 0 & v_\eta \end{bmatrix}. \tag{19}$$

The linear pressure-to-density relationship $p = c_0^2 \rho$ is used here. For a characteristic formulation, Eq. (18) has to be diagonalized. This can be done by the application of the transformation matrix S on A. The matrix S consists of the eigenvectors of the matrix A

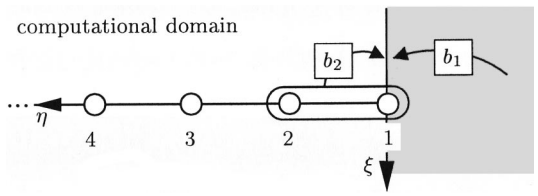


FIG. 4. Boundary conditions in characteristic form. At the boundary the acoustic field is split into an incoming wave b_1 and an outgoing wave b_2, which is calculated from inner grid points 1 and 2.

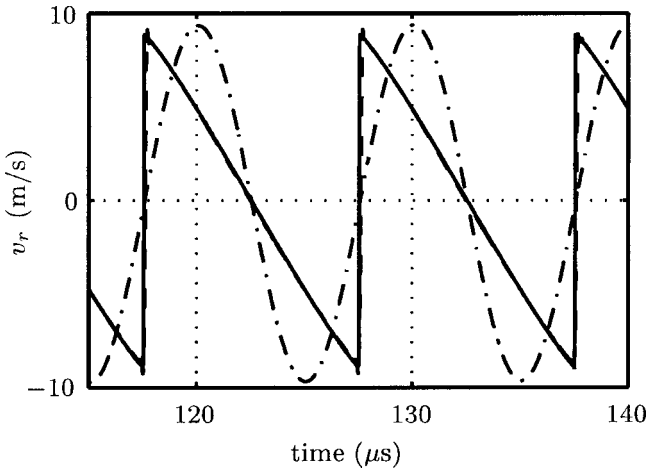


FIG. 5. Comparison of the nonlinear (- -) and linear (-··-) simulation results with the analytical (——) solution.

$$S = \begin{bmatrix} \dfrac{\rho + \rho_0}{c_0} & -\dfrac{\rho + \rho_0}{c_0} & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{20}$$

In characteristic form, one finally obtains

$$S^{-1} \frac{\partial \mathbf{w}}{\partial t} + \underbrace{\Lambda S^{-1} \frac{\partial \mathbf{w}}{\partial \eta}}_{\mathbf{L}} + S^{-1} \mathbf{F}_{\text{trans}} = 0, \tag{21}$$

where the diagonal matrix $\Lambda = S^{-1} A S = \text{diag}(v_\eta + c_0, v_\eta - c_0, v_\eta)$ is composed by the eigenvalues of A. The vector $\mathbf{L}$ written in components gives

$$\begin{aligned} &\mathbf{L} = [b_1 \Lambda_1, b_2 \Lambda_2, b_3 \Lambda_3]^T, \\ &b_1 = \frac{1}{2} \frac{\partial v_\eta}{\partial \eta} + \frac{1}{2} \frac{c_0}{(\rho_0 + \rho)} \frac{\partial \rho}{\partial \eta}, \\ &b_2 = \frac{1}{2} \frac{\partial v_\eta}{\partial \eta} - \frac{1}{2} \frac{c_0}{(\rho_0 + \rho)} \frac{\partial \rho}{\partial \eta}, \quad b_3 = \frac{\partial v_\xi}{\partial \eta}. \end{aligned} \tag{22}$$

Here, b_1 and b_2 are the acoustic incoming and outgoing waves, respectively which belong to the eigenvalues $v_\eta + c_0$ and $v_\eta - c_0$; b_3 describes the vorticity wave. The outgoing wave b_2 is fully determined by the known values at the inner grid point. The incoming wave b_1 has to be determined to fulfill the boundary condition at the boundary. Fig-

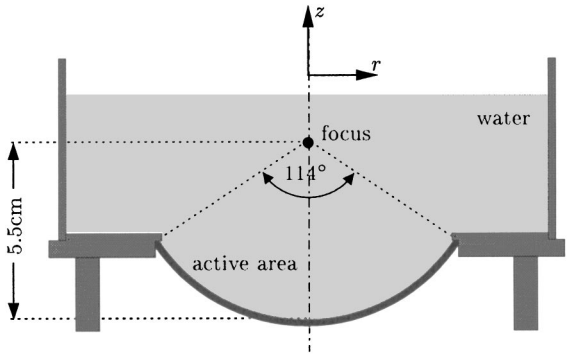


FIG. 6. Geometry of the self-focusing piezoelectric transducer.

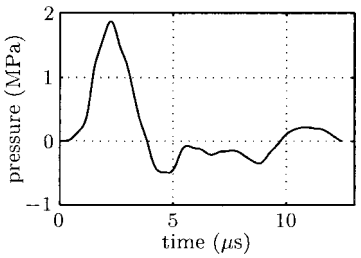


FIG. 7. Measured initial source pressure signal for the self-focusing transducer (Fig. 6).

ure 4 depicts the characteristic boundary treatment, which is applied only to the last grid point. The outgoing wave b_2 is calculated using one-sided differentiation stencils. In the points 2 and 3 the normal conservation formulation is used with symmetric stencils of reduced order. To incorporate the boundary treatment into the regular integration algorithm, as described above, the characteristic form is partially transformed back to the quasilinear form Eq. (23), but expressed in the wave components b_i,

$$\frac{\partial \mathbf{w}}{\partial t} + \underbrace{S\mathbf{L} + \mathbf{F}_{\text{trans}}}_{\mathbf{H}_{i,j}^n} = 0. \tag{23}$$

So, only the calculation of the differentiated flux $\mathbf{H}_{i,j}^n$ is changed, whereas the integration step remains the same throughout the procedure.

At the absorbing boundary, no wave enters the computational domain. This means that the time derivative of the first equation in Eq. (21) must be zero, $[S^{-1}\partial \mathbf{w}/\partial t]_1 = 0$. Solving the first equation in Eq. (21) for b_1, one obtains

$$b_{1,\text{abs}} = -\frac{1}{2}\frac{(\rho_0+\rho)c_0 F_{\text{trans},1} + F_{\text{trans},2}}{(\rho_0+\rho)c_0 \Lambda_1}. \tag{24}$$

At an impedance boundary the reflected wave b_1 depends on the outgoing wave b_2 and on the acoustic impedance Z_b at that boundary.[33] Differentiating the impedance

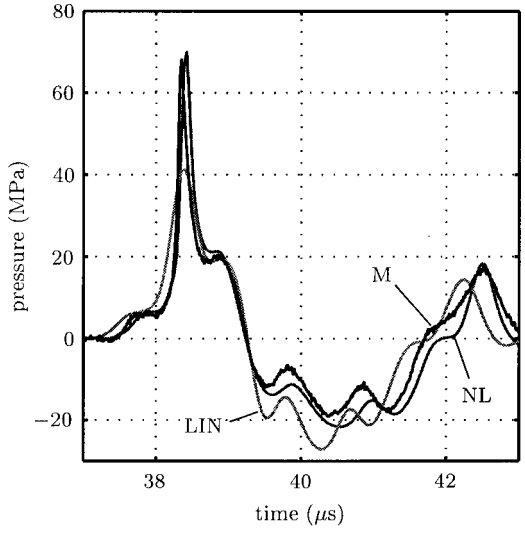


FIG. 8. Comparison of nonlinear calculated (NL), linear calculated (LIN), and measured (M) pressure pulses in the acoustic focus of the self-focusing transducer (Fig. 6).

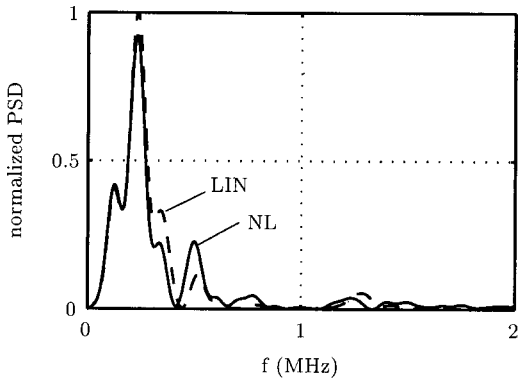


FIG. 9. Power spectral densities of focal pressures normalized to the maximum of the linear calculation.

relation at the boundary $Z_b = p/v_\eta$ and using the linear state equation $p = c_0^2 \rho$, one obtains

$$c_0^2 \frac{\partial \rho}{\partial t} - v_\eta \frac{\partial Z_b}{\partial t} - Z_b \frac{\partial v_\eta}{\partial t} = 0. \tag{25}$$

After substitution of $\partial\rho/\partial t$ and $\partial v_\eta/\partial t$ in Eq. (25) with the first and second component of Eq. (23), one can solve the equation for b_1 at the impedance boundary

$$b_{1,\text{imp}} = \frac{((\rho_0+\rho)c_0 - Z_b)\Lambda_2 b_2}{((\rho_0+\rho)c_0 + Z_b)\Lambda_1} - \frac{Z_b F_{\text{trans},1} + F_{\text{trans},2} + v_\eta \dfrac{\partial Z_b}{\partial t}}{((\rho_0+\rho)c_0 + Z_b)\Lambda_1}. \tag{26}$$

At the physical existing source, the incoming wave b_1 depends on the given source pressure $p_s(t)$ and on the outgoing wave b_2, which is reflected at the source boundary according to the acoustic impedance Z_b of the boundary material. Hence, $b_{1,\text{src}}$ can be written as a superposition of an incoming transparent source wave $b_{1,\text{src-tran}}$ (Ref. 30) and a reflected wave $b_{1,\text{imp}}$,

$$b_{1,\text{src}} = b_{1,\text{src-tran}} + b_{1,\text{imp}}. \tag{27}$$

The addend $b_{1,\text{imp}}$ is given in Eq. (26). To obtain $b_{1,\text{src-tran}}$ the first component of Eq. (21) is solved for the unknown b_1. After replacing the time derivatives of ρ and v_η by the corresponding source quantities $\partial\rho_s/\partial t = (1/c_0^2)\partial p_s/\partial t$ and $\partial v_s/\partial t = \pm 1/(c_0\rho_0)\partial p_s/\partial t$, one gets

$$b_{1,\text{src-tran}} = -\frac{1}{2}\frac{c_0}{(\rho_0+\rho)\Lambda_1}\frac{\partial \rho_s}{\partial t} - \frac{1}{2\Lambda_1}\frac{\partial v_s}{\partial t} - \frac{1}{2}\frac{c_0(\rho_0+\rho)F_{\text{trans},1} + F_{\text{trans},2}}{c_0(\rho_0+\rho)\Lambda_1}. \tag{28}$$

IV. VERIFICATION

The verification of the nonlinear US propagation model is performed by comparison of the numerical calculations with an analytical one-dimensional solution and with measurements in two different focusing transducers. As a first practical example, a self-focusing piezoelectric transducer (Fig. 6) is investigated. In a second example, a reflector fo-

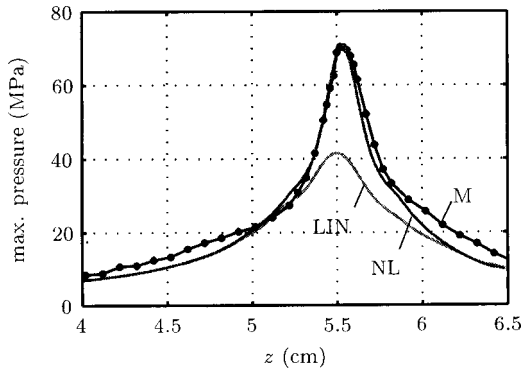


FIG. 10. Comparison of nonlinear calculated (NL), linear calculated (LIN) and measured (M) axial distribution of the maximal pressure amplitudes for the self-focusing transducer (Fig. 6).

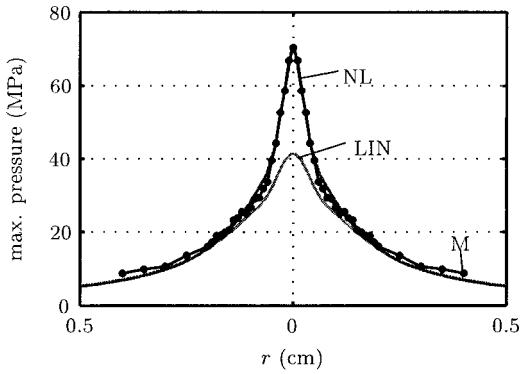


FIG. 11. Comparison of nonlinear calculated (NL), linear calculated (LIN), and measured (M) lateral distribution of the maximal pressure amplitudes in the focal plane for the self-focusing transducer (Fig. 6).

cusing electromagnetic transducer (Fig. 12) is validated. The measurements are performed in degassed water with a fiber-optical probe hydrophone,[34] which is an accepted measurement device for shock wave measurements in lithotripsy.[35]

A. Comparison with analytical solution

Blackstock's one-dimensional analytical result for a cylindrically diverging sinusoidal wave of finite amplitude[36] is applied to demonstrate the correct nonlinear propagation behavior of the presented numerical model. The analytical solution provides a first-order approximation of the exact shock-propagation speed. The comparison is performed with a 100-kHz sine wave, radiated from an infinite cylinder with a radius of $r_0=4$ cm into a lossless fluid with the acoustic properties of water at $T=22\,^{\circ}\mathrm{C}$ ($c_0=1487$ m/s, $B/A=4.9$).[37] At the cylinder's surface the radial component of the velocity is $v_r(r_0)=21.67$ m/s (corresponding to a shock distance of 10 cm). To limit the computational domain, symmetric and absorbing boundary conditions are implemented, as described in Sec. III. The wave profiles at $r=20$ cm of the analytical, the nonlinear numerical (with $\Delta x=50$ μm), and the linear numerical (with $\Delta x=300$ μm) results are shown in Fig. 5. The comparison with the linear result shows an anomalous dissipation in the analytical as well as in the numerical solution. The excellent agreements of the zeros even in the shock locations demonstrate a correct reproduction of the shock speed in a first-order approximation.

B. Measurements and calculations in the self-focusing transducer

Figure 6 shows the arrangement of the self-focusing transducer. The transducer has a focal distance of 55 mm and is filled with degassed water at a temperature of $T=20\,^{\circ}\mathrm{C}$ ($\rho_0=998$ kg/m^3, $c_0=1482$ m/s, $B/A=4.9$).[37] The measured initial pressure signal at the surface of the transducer is shown in Fig. 7.

Two different calculations are performed. For the first calculation the nonlinear (NL) numerical model, Eqs. (12) and (13), are used, with a spatial discretization of $\Delta x=50$ μm and a Courant number $\gamma=0.125$. For comparison a second calculation (LIN) is performed in the same arrangement, neglecting the nonlinear terms in the acoustic variables. Here, no steepening occurs and therefore a discretization $\Delta x=75$ μm is sufficient. The calculation of the nonlinear problem using an HP C3000 requires 320 MB of RAM and a computational time of about 25 h.

Figure 8 shows the comparison between the measured (M) and the calculated (NL, LIN) pressure–time waveforms in the acoustic focus. The result of the NL calculation follows the measured waveform closely, whereas the linear calculation severely underestimates the measured maximal amplitude and overestimates the maximal negative pressure. In Fig. 9 the power spectral densities (PSD) of the linear (LIN) and nonlinear (NL) calculated signals of Fig. 8 are shown, normalized to the maximum value of the linear one. The differences demonstrate the shift from lower to higher frequency components due to nonlinear effects. In Fig. 10 the distribution of the maximal pressure amplitudes on the z axis is presented and Fig. 11 shows the maximal pressure amplitudes in their lateral distribution in the focal plane. The therapeutically important features, namely the maximal and the minimal pressure amplitudes, the nonlinear steepened shock front or the -6-dB area of the focal zone, are especially well matched. This comparison clearly points out that

TABLE I. Quantitative comparison of the measured and predicted field parameters for the self focusing transducer (Fig. 6).

		Calculation			
	Meas.	Nonlinear		Linear	
$p_{\max}$ (MPa)	68.1	70.6	(+3.6%)	41.5	(−39.0%)
$p_{\min}$ (MPa)	−19.2	−21.6	(−12.5%)	−27.1	(−41.1%)
t_{FWHM} (μs)	0.186	0.199	(+7.0%)	0.501	(+169.3%)
−6 dB in z (mm)	4.66	4.49	(−3.7%)	9.19	(+97.2%)
−6 dB in r (mm)	1.15	1.28	(+11.3%)	2.78	(+141.7%)

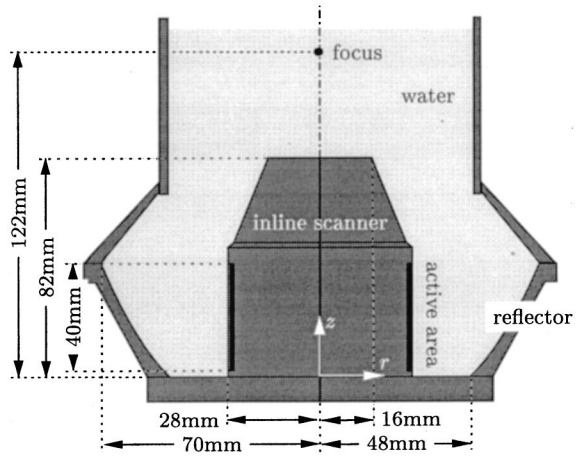


FIG. 12. Geometry of the reflector focusing electromagnetic transducer. The curvature of the paraboloid reflector is given by $r(z) = -2.785(z-0.122)^2 + 0.0898$.

nonlinear steepening effects are important in therapeutic US devices. From Figs. 10 and 11 one can deduce the spatial region, where this nonlinear effects drastically alter the pressure distribution. Table I quantitatively summarizes the characteristic parameters of the pressure distribution for the self-focusing transducer.

C. Measurements and calculations in the reflector focusing transducer

In contrast to methods based on a parabolic approximation of the nonlinear wave equations like the KZK equation, the numerical model presented here is also applicable to calculate nonlinear wave propagation in reflector focusing transducers. Figure 12 shows the electromagnetic reflector focusing transducer investigated here. The initial pressure signal is generated by a cylindrical electromagnetic source.[38] Figure 13 depicts the measured pressure pulse at the radiating area. The initial cylindrical wave is transformed by the parabolic brass reflector into a converging spherical wave.

Measurements are performed with a fiber-optic hydrophone in degassed water at a temperature of $T = 22\,^{\circ}\mathrm{C}$. For the simulations following material parameters[37] for water $\rho_0 = 998\ \mathrm{kg/m^3}$, $c_0 = 1487$ m/s, $B/A = 4.9$ are used as before. Based on a spatial-temporal transformation of the equations, the discretization of the computational domain is adaptively refined as the spherical pressure wave converges to the focus. Hence, the FDTD algorithm always resolves the developing weak shock very well with minimal additional numerical

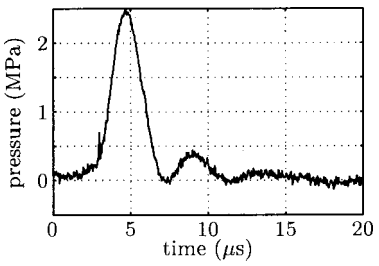


FIG. 13. Measured initial pressure signal for the reflector focusing transducer (Fig. 12).

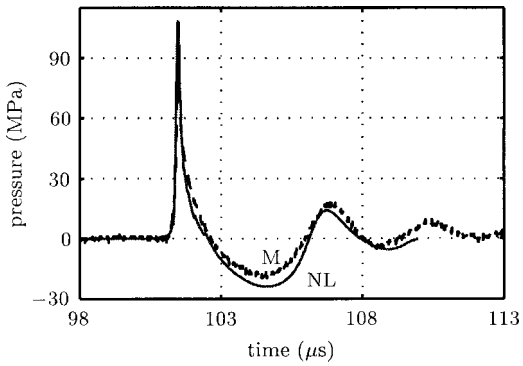


FIG. 14. Comparison of nonlinear calculated (NL) and measured (M) pressure pulses in the acoustic focus of the reflector focusing transducer (Fig. 12).

costs. For more details on this adaptive zooming see Refs. 25 and 39. The calculations are performed using the nonlinear model Eqs. (12) and (13), with high diligence to the FDTD implementation of the curved boundary at the brass reflector.

In Figs. 14 and 15 the measured (M) and calculated (NL) pressure–time waveforms are compared. Figure 14 shows the pressure–time waveforms in the acoustic focus and Fig. 15 those in the focal plane 14 mm off axis. Both diagrams demonstrate an excellent agreement of the predicted pressure–time curves with the measured ones, regarding the shock front, the maximal pressure amplitude in the focus, and the waveform of the signal off-axis.

V. CONCLUSION

Nonlinear models are required to calculate therapeutically applied ultrasound. Linear US propagation models underestimate not only the maximal pressure amplitudes in the focus but also the pressure rise at the front of the wave. Comparisons to measurements clearly demonstrate that the nonlinear FDTD model is able to predict the pressure distribution in different therapeutic ultrasound transducers very accurately. The advantage of the presented model lies in the combination of different aspects: first, a nonlinear acoustic approximation in conservation form of the full nonlinear hydrodynamic equations is used to include weak shocks. Second, the high-order FDTD algorithm combined with the dispersion relation preserving (DRP) method for coefficient

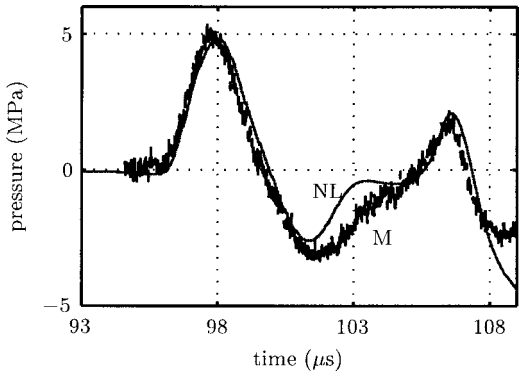


FIG. 15. Comparison of nonlinear calculated (NL) and measured (M) pressure pulses in the focal plane at $r = 14$ mm off axis of the reflector focusing transducer (Fig. 12).

determination provides an optimal numerical treatment of smooth acoustic solutions. And finally, the specially adapted high-order implementation of the ideal reflecting boundary conditions completes the model.

Between the simulation and experimental results an agreement of about 10% is observed. Therefore, it can be concluded that the presented method provides field predictions with high accuracy. It can be used to investigate the behavior of new therapeutic devices or to optimize existing therapy systems. Beside a reduction in development time and costs, it enables a deeper insight into the field characteristics that point-by-point measurements can provide.

An extension for US fields in attenuating tissue layers, which typically show attenuation according to a frequency power law, will be the topic of a subsequent paper. With this extension, studies of complex effects like the enhanced absorption due to the counteracting processes of nonlinear steepening and frequency-dependent tissue absorption are possible. Also, further investigations on the development and influence of the tissue temperature on the US propagation will be possible. Such considerations are important in HIFU treatment of biological tissues, where in vivo measurements are still a very difficult task.

ACKNOWLEDGMENTS

The authors would like to thank the Richard Wolf GmbH (D-75434 Knittlingen, Germany), who kindly provided the self-focusing piezoelectric transducer and the Storz Medical AG (CH-8280 Kreuzlingen, Switzerland), who made the parabolic reflector focusing, electromagnetic transducer available to us. Parts of this work were supported by the Deutsche Forschungsgemeinschaft (DFG) under Grant No. Wi 1044/3.

APPENDIX A: COEFFICIENTS OF THE APPLIED FDTD ALGORITHM

The applied coefficients for the numerical implementation of the FDTD algorithm from Sec. III B are listed in Refs. 25 and 40. The following coefficients b_r are applied for the numerical integration step of Eq. (12):

$$\begin{aligned} b_0 &= +2.302\,558\,088\,838, \\ b_1 &= -2.491\,007\,599\,848, \\ b_2 &= +1.574\,340\,933\,128, \\ b_3 &= -0.385\,891\,422\,172. \end{aligned} \tag{A1}$$

The numerical differentiation of Eq. (13) works with a symmetric, seven-point stencil with the weights

$$\begin{aligned} a_3 &= +0.020\,843\,142\,770 = -a_{-3}, \\ a_2 &= -0.166\,705\,904\,414 = -a_{-2}, \\ a_1 &= +0.770\,882\,380\,518 = -a_{-1}, \\ a_0 &= 0. \end{aligned} \tag{A2}$$

Finally, for the artificial numerical viscosity term in Eq. (13), the coefficients of the underlying filter are

$$\begin{aligned} c_3 &= -0.014\,281\,184\,692 = c_{-3}, \\ c_2 &= +0.086\,150\,669\,577 = c_{-2}, \\ c_1 &= -0.235\,718\,815\,308 = c_{-1}, \\ c_0 &= +0.327\,698\,660\,846. \end{aligned} \tag{A3}$$

[1] M. Delius, "Lithotripsy," Ultrasound Med. Biol. **26**(1), 55–58 (2000).
[2] W. Siebert and M. Buch, *Extracorporeal Shock Waves in Orthopaedics* (Springer, Berlin, 1998).
[3] C. G. Chaussy and S. Thüroff, "High-intensity focused ultrasound in localized prostate cancer," J. Endourol **14**(3), 293–299 (2000).
[4] G. R. ter Haar, "Intervention and therapy," Ultrasound Med. Biol. **26**(1), 51–54 (2000).
[5] S. Vaezy, R. Martin, B. Goldman, E. Chi, W. Chandler, P. Kaczkowski, and L. Crum, "Biological mechanisms of acoustically-induced hemostasis," in IEEE Ultrasonics Symposium Proceedings, Caesars Tahoe, Nevada, Vol. 2, pp. 1401–1404, 1999.
[6] V. P. Kuznetsov, "Equations of nonlinear acoustics," Sov. Phys. Acoust. **16**(4), 467–470 (1971).
[7] *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, London, 1998), Chaps. 3, 4, 11.
[8] I. M. Hallaj and R. O. Cleveland, "FDTD simulation of finite-amplitude pressure and temperature fields for biomedical ultrasound," J. Acoust. Soc. Am. **105**, L7 (1999).
[9] G. Wojcik, J. Mould, F. Lizzi, N. Abboud, M. Ostromogilsky, and D. Vaughan, "Nonlinear modeling of therapeutic ultrasound," in IEEE Ultrasonics Symposium Proceedings, pp. 1617–1622, 1995.
[10] T. Yano and Y. Inoue, "Strongly nonlinear waves and streaming in the near field of a circular piston," J. Acoust. Soc. Am. **99**, 3353–3372 (1996).
[11] Y.-S. Lee and M. F. Hamilton, "Time-domain modeling of pulsed finite-amplitude sound beams," J. Acoust. Soc. Am. **97**, 906–917 (1995).
[12] R. O. Cleveland, M. F. Hamilton, and D. T. Blackstock, "Time-domain modeling of finite-amplitude sound beams in relaxing fluids," J. Acoust. Soc. Am. **99**, 3312–3318 (1996).
[13] J. Bernsten and B. Ystad, "Numerical solution of parabolic equations for strongly curved focusing sources," Acta Acust. (Beijing) **82**, 698–706 (1996).
[14] J. N. Tjøtta and S. Tjøtta, "Propagation and interaction of two collinear finite amplitude sound beams," J. Acoust. Soc. Am. **88**, 2859–2870 (1990).
[15] V. F. Humphrey, "Nonlinear propagation in ultrasonic fields: Measurements, modeling and harmonic imaging," Ultrasonics **38**, 267–272 (2000).
[16] O. V. Rudenko and A. K. Sukhorukova, "Nonlinear sawtooth waves in an inhomogeneous medium," Sov. Phys. Acoust. **37**(4), 392–395 (1992).
[17] M. A. Averkiou and R. O. Cleveland, "Modeling of an electrohydraulic lithotripter with the KZK equation," J. Acoust. Soc. Am. **106**, 102–112 (1999).
[18] P. T. Christopher and K. J. Parker, "New approaches to nonlinear diffractive field propagation," J. Acoust. Soc. Am. **90**, 488–499 (1991).
[19] T. Christopher, "Modeling the Dornier HM3 lithotripter," J. Acoust. Soc. Am. **96**, 3088–3095 (1994).
[20] J. Tavakkoli, D. Cathignol, and R. Souchon, "Modeling of pulsed finite-amplitude focused sound beams in time domain," J. Acoust. Soc. Am. **104**, 2061–2072 (1998).
[21] L. D. Landau and E. M. Lifschitz, *Hydrodynamik*, 5th ed. (Akademie GmbH, Berlin, 1991), Chaps. 1, 9.
[22] A. D. Pierce, *Acoustics—An Introduction to its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY, 1991), Chaps. 1, 11.
[23] V. A. Šutilov, *Physik des Ultraschalls* (Physics of Ultrasound) (Springer, Wien, 1984), Chap. 4.
[24] W. Eisenmenger, "Experimentelle Bestimmung der Stossfrontdicke aus dem akustischen Frequenzspektrum elektromagnetisch erzeugter Stosswellen in Flüssigkeiten bei einem Stossdruckbereich von 10 atm bis 100 atm (Experimental determination of the shock front thickness from the acoustic frequency spectrum of electromagnetic generated shockwaves in fluids at pressures from 10 to 100 atm.)," Acustica **14**(4), 187–204 (1964).
[25] E. Steiger, "Modellierung der Ausbreitung in extrakorporalen Therapien eingesetzter Ultraschallimpulse hoher Intensität (Modeling of the propa-

gation of ultrasonic impulses used in extracorporal therapies),'' Ph.D. thesis, Fakultät für Elektrotechnik und Informationstechnik, Universität Karlsruhe, Karlsruhe, Germany, 1998.
[26] R. J. Le Veque, *Numerical Methods for Conservation Laws*, 2nd ed. (Birkhäuser, Basel, 1992), Chaps. 1, 12.
[27] V. Sparrow and R. Raspet, ''A numerical method for general finite amplitude wave propagation in two dimensions and its application to spark pulses,'' J. Acoust. Soc. Am. **90**, 2683–2691 (1991).
[28] C. K. Tam and J. C. Webb, ''Dispersion-relation-preserving finite difference schemes for computational acoustics,'' J. Comput. Phys. **107**, 262–281 (1993).
[29] E. Steiger, ''Scattering of focused strong pressure pulses by spheroids,'' in Proceedings of the International Symposium on Hydroacoustics and Ultrasonics, pp. 97–102, 1997.
[30] K. W. Thompson, ''Time-dependent boundary conditions for hyperbolic systems,'' J. Comput. Phys. **68**, 1–24 (1987).
[31] K. W. Thompson, ''Time-dependent boundary conditions for hyperbolic systems, II.'' J. Comput. Phys. **89**, 439–461 (1990).
[32] C.-J. Hwang and D.-J. Lee, ''Transparent acoustic source condition applied to the Euler equations,'' Am. Inst. Aeronaut. Astronaut. **33**(9), 1736–1738 (1995).
[33] W. R. Watson and M. K. Myers, ''Inflow-outflow boundary conditions for two-dimensional acoustic waves in channels with flow,'' Am. Inst. Aeronaut. Astronaut. **29**(9), 1383–1389 (1991).
[34] W. Eisenmenger, ''Fiber-optic probe hydrophone for ultrasonic and shock-wave measurements in water,'' Ultrasonics **31**, 267–273 (1993).
[35] IEC 61846. Ultrasonics—Pressure Pulse Lithotripters—Characteristics of Fields, International standard, International Electrotechnical Commission, 1998.
[36] D. T. Blackstock, ''Connection between the Fay and Fubini solutions for plane sound waves of finite amplitude,'' J. Acoust. Soc. Am. **39**, 1019–1026 (1966).
[37] F. A. Duck, *Physical Properties of Tissue* (Academic, London, 1990).
[38] E. Steiger, ''Full wave modeling of lithotripter fields,'' J. Acoust. Soc. Am. **103**, 2795–2796(A) (1998).
[39] E. Steiger, ''FDTD-modeling of propagation of high energy sound pulses in lithotripter–tissue arrangements,'' in *IEEE International Symposium Proceedings*, Toronto, Canada, Vol. 2, pp. 1361–1364, 1997.
[40] C. K. Tam, J. C. Webb, and Z. Dong, ''A study of the short wave components in computational acoustics,'' J. Comp. Acoust. **1**(1), 1–30 (1993).

Directional dependence of nonlinear surface acoustic waves in the (001) plane of cubic crystals

R. E. Kumon[a] and M. F. Hamilton
Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas 78712-1063

(Received 16 August 2001; revised 19 December 2001; accepted 7 January 2002)

Spectral evolution equations are used to perform analytical and numerical studies of nonlinear surface acoustic waves in the (001) plane of a variety of nonpiezoelectric cubic crystals. The basic theory underlying the model equations is outlined, and quasilinear solutions of the equations are presented. Expressions are also developed for a characteristic length scale for nonlinear distortion and a nonlinearity coefficient. A time-domain equation corresponding to the spectral equations is derived. Numerical calculations based on measured second- and third-order elastic constants taken from the literature are performed to predict the evolution of initially monofrequency surface waves. Nonlinearity matrix elements that indicate the coupling strength of harmonic interactions are shown to provide a useful tool for characterizing waveform distortion. The formation of compression or rarefaction shocks can be strongly dependent on the direction of propagation, and harmonic generation is suppressed or increased in certain directions. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1455023]

PACS numbers: 43.25.Fe, 43.25.Dc [ANN]

I. INTRODUCTION

Unlike the theory for nonlinear Rayleigh waves in isotropic media, the theory for nonlinear surface acoustic waves (SAWs) in crystalline media predicts a wide diversity of effects that depend on the elastic and symmetry parameters of the materials, orientation of the surface cut with respect to the crystalline axes, and direction of propagation. Simulations based on numerical solution of spectral evolution equations introduced by Hamilton, Il'inskii, and Zabolotskaya[1] are presented for eight different crystals (KCl, NaCl, SrF_2, BaF_2, Si, Ge, Ni, Cu) over a full range of propagation directions in the (001) surface cut (results for additional cubic crystalline materials and symmetries are presented elsewhere[2]). The nonlinearity matrix used to describe harmonic interactions generally has complex-valued elements with nonuniform phase for an arbitrary surface cut. However, because of the symmetries in the (001) surface cut of cubic crystals, the nonlinearity matrix for this plane can always be written in a real-valued form, similar to that for the case of nonlinear Rayleigh waves. While the qualitative nature of the waveform distortion of surface waves in the (001) plane is similar to that of nonlinear Rayleigh waves in some cases, the simulations show that the distortion may depend strongly on the direction of propagation. For example, compression shocks may form in some directions but rarefaction shocks in others. In addition, there exist particular directions where the distortion differs completely from nonlinear Rayleigh waves because of suppression or enhancement of generation of one or more harmonics. Measurements in crystalline silicon corroborate several of the predicted results.[3]

While extensive work has been performed to investigate nonlinear surface waves,[4,5] most of it has been limited to isotropic media. Researchers first began to develop theories to predict the propagation of nonlinear SAWs in nonpiezoelectric crystals in the mid-1980s. Planat[6] developed a theory for an elastic solid with general anisotropy based on a multiple scales approach but presented results for only a limited number of harmonics. Another multiple scales theory with general anisotropy by Lardner[7] was employed by Lardner and Tupholme[8] to investigate the properties of nonlinear surface waves in cubic crystals, but the reported results were limited to propagation along a crystalline axis in a plane of symmetry. Results were provided as tables of coefficients describing the growth and decay rates of the fundamental, second, and third harmonics for 35 cubic crystals or crystalline alloys, as well as for the first three harmonic propagation curves for the specific case of magnesium oxide. Independently, Parker[9] developed a theory for nonpiezoelectric, anisotropic media that avoids some of the complications and limitations of the multiple scales approach by introducing a reference frame moving at the linear wave speed to derive spectral evolution equations. Results were given only in terms of waveforms for an isotropic solid (although numerical results were later presented for a piezoelectric crystal[10]).

During the mid-1990s, Hamilton *et al.*[1,11] extended the isotropic theory of Zabolotskaya for nonlinear Rayleigh waves[12,13] to include nonpiezoelectric, anisotropic media and presented results for potassium chloride in selected propagation directions. Piezoelectric effects were subsequently included in this theory.[14] Gusev *et al.*[15] also developed a theory for nonlinear SAWs in anisotropic media, but numerical calculations demonstrating their results have not been reported. Their evolution equations are given in the time domain and differ from those of Parker[9] and Hamilton *et al.*[1] Further discussion of the approach used by Gusev *et al.*[15] is provided by Meegan *et al.*[16]

In recent experiments, very-high-amplitude SAWs

[a]Current address: National Institute of Standards and Technology, 325 Broadway, Mail Stop 853, Boulder, CO 80305-3328. Electronic mail: kumon@mailaps.org

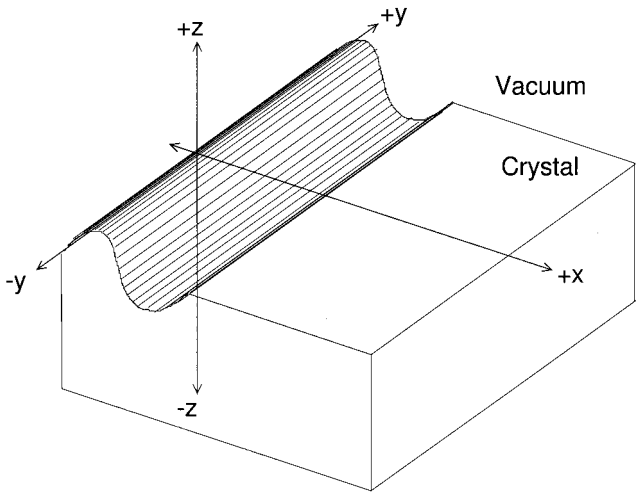


FIG. 1. Coordinate system for plane surface wave propagation. Note that the positive z axis points out of the crystal.

(strains up to 0.01) have been generated photoelastically in fused quartz,[17–19] polycrystalline aluminum and copper,[20] and crystalline silicon[17,21] via pulsed laser excitation. Calculations based on the theories of Zabolotskaya[12] for isotropic media and Hamilton *et al.*[1] for crystals agree well with the measured results.[3,22,23]

II. NONLINEAR THEORY

A. Model equations

A plane surface wave is assumed to propagate in the x direction in an anisotropic half-space $z \leqslant 0$ (see Fig. 1). The velocity components of the surface wave are taken to have the form

$$v_j(x,z,\tau)=\sum_{n=1}^{\infty} v_n(x)u_{nj}(z)e^{-in\omega\tau}+\text{c.c.}, \tag{1}$$

where $j=x,y,z$ is the coordinate index, ω is the fundamental angular frequency in the expansion, $\tau=t-x/c$ is the retarded time, and c is the linear wave speed in the direction of propagation. The depth dependence is given by the functions

$$u_{nj}(z)=\sum_{s=1}^{3} \beta_j^{(s)} e^{ink\zeta_s z}, \tag{2}$$

where ζ_s and $\beta_j^{(s)}$ are computed from the eigenvalues and eigenvectors, respectively, of the linear boundary value problem corresponding to the stress-free surface,[24] and $k=\omega/c$ is the fundamental wave number. The coupled spectral evolution equations for the slowly varying amplitudes $v_n(x)$ are given by[1]

$$\frac{dv_n}{dx}+\alpha_n v_n=\frac{n^2\omega}{2\rho c^4}\left(\sum_{m=1}^{n-1} S_{m,n-m}v_m v_{n-m} - 2\sum_{m=n+1}^{\infty} S^*_{n,m-n}v_m v^*_{m-n}\right), \tag{3}$$

where α_n is the absorption coefficient for the nth harmonic, ρ is the density, and S_{lm} is the nonlinearity matrix. Equations (3) are written in terms of Lagrangian coordinates, an approach which is convenient for describing the surface boundary conditions. The nonlinearity matrix is given by

$$S_{lm}=\sum_{s_1,s_2,s_3=1}^{3} \frac{F_{s_1s_2s_3}}{l\zeta_{s_1}+m\zeta_{s_2}-(l+m)\zeta^*_{s_3}}, \tag{4}$$

where

$$F_{s_1s_2s_3}=\tfrac{1}{2}d'_{ijklmn}\beta_i^{(s_1)}\beta_k^{(s_2)}[\beta_m^{(s_3)}]^* l_j^{(s_1)} l_l^{(s_2)}[l_n^{(s_3)}]^*, \tag{5}$$

$$d'_{ijklmn}=d_{ijklmn}+c_{ijln}\delta_{km}+c_{jnkl}\delta_{im}+c_{jlmn}\delta_{ik}, \tag{6}$$

$\mathbf{l}^{(s)}=(1,0,\zeta_s)$, and c_{ijkl} and d_{ijklmn} are the second-and third-order elastic constants transformed to the chosen reference frame. (Elastic constants are typically listed with respect to the coordinate system defined by the crystalline axes.)

The following source conditions are assumed at $x=0$:

$$v_1=v_0, \quad v_{n>1}=0, \tag{7}$$

where v_0 is taken to be a real-valued characteristic amplitude. Substitution of Eqs. (7) into (1) and (2) yields the following source velocity components at the surface:

$$v_j(0,0,t)=2v_0|B_j|\cos(\omega t-\phi_j), \tag{8}$$

where

$$B_j=|B_j|e^{i\phi_j}=\sum_{s=1}^{3}\beta_j^{(s)}. \tag{9}$$

Note that only the relative phases of B_j are determined by the theory[1]—the absolute phases may be chosen as is convenient. While all simulations considered in this paper are for a sinusoidal source, any source spectrum can be employed via Fourier decomposition. For the simulations shown here, the absolute phases are chosen such that B_1 lies along the negative imaginary axis. In the limit of an isotropic solid we may choose $v_y=0$, and Eqs. (8) reduce to

$$\begin{aligned} v_x(0,0,t)&=-2v_0|B_1|\sin\omega t,\\ v_z(0,0,t)&=2v_0|B_3|\cos\omega t. \end{aligned} \tag{10}$$

B. Second harmonic and nonlinearity coefficient

To gain insight into the nature of the lowest-order harmonic generation, consider Eqs. (3) with ∞ in the second summation replaced by 2:

$$\frac{dv_1}{dx}+\alpha_1 v_1=0, \quad \frac{dv_2}{dx}+\alpha_2 v_2=\frac{2S_{11}\omega}{\rho c^4}v_1^2, \tag{11}$$

and the corresponding solutions satisfying Eqs. (7):

$$v_1=v_0e^{-\alpha_1 x}, \quad v_2=\frac{2S_{11}\omega v_0^2}{\rho c^4}\left(\frac{e^{-2\alpha_1 x}-e^{-\alpha_2 x}}{\alpha_2-2\alpha_1}\right). \tag{12}$$

Note that S_{11} is generally complex-valued, and thus $S_{11}/|S_{11}|=e^{i\psi_{11}}$ with $0\leqslant\psi_{11}\leqslant 2\pi$. In the limit of no absorption Eqs. (12) reduce to

$$v_1/v_0=1, \quad v_2/v_0=(2S_{11}\omega v_0^2/\rho c^4)x. \tag{13}$$

The similarity of Eqs. (13) with the corresponding solutions for sound waves in lossless fluids makes it possible to propose an estimate of the shock formation distance. Let $\overline{x_f}$

be the shock formation distance for a finite-amplitude sound wave radiated at angular frequency ω and velocity amplitude v_0 in a lossless fluid. The quasilinear solution for the second harmonic component is[25] $|v_2|/v_0 = x/2\overline{x_f}$. Comparison with Eqs. (13) shows that an estimate of the shock formation distance x_0 for the SAW is thus

$$x_0 = \frac{\rho c^4}{4|S_{11}|\omega v_0}, \tag{14}$$

which depends on the third-order elastic constants via the nonlinearity matrix element S_{11}.

Equation (14) can be somewhat problematic when applied to crystalline media. Harmonic generation may occur while shock formation does not, and therefore the characteristic distance given by Eq. (14) loses meaning in those cases. For example,[1] near the ⟨100⟩ direction in the (001) cut of KCl, the transfer of energy from the fundamental to higher harmonics is so weak that shocks never form (see Sec. III B 3). Despite the fact that Eq. (14) is not applicable in all cases, it does provide at least a first-order estimate of the shock formation distance in many cases of interest and can still be useful in this respect.

An expression for a nonlinearity coefficient similar to that of fluids may also be derived. For a finite-amplitude wave radiated at angular frequency ω and velocity amplitude v_0 propagating in such a medium, the shock formation distance $\overline{x_f}$ is related[25] to the coefficient of nonlinearity β_f by $\overline{x_f} = c^2/|\beta_f|v_0\omega$. From Eq. (14), a comparable SAW nonlinearity coefficient is

$$\beta = -4S_{11}/\rho c^2. \tag{15}$$

In general, β is complex-valued, although cases do exist where it is strictly real-valued. The negative sign is introduced in Eq. (15) so that the sign of the nonlinearity coefficient is consistent with the corresponding theory for isotropic media.[26]

C. Time-domain evolution equation

As for isotropic solids,[12] a time-domain equation for the evolution of nonlinear SAWs on the surface of crystals ($z=0$) may be derived from the frequency-domain evolution equations, Eqs. (3), without the absorption term. Equations (3) may be rewritten with the summation on the right side performed over all indices, both positive and negative:

$$\frac{dv_n}{dx} = \frac{n^2\omega}{2\rho c^4} \sum_{l+m=n} \frac{lm}{|lm|} S_{lm} v_l v_m. \tag{16}$$

From Eqs. (1) at $z=0$ we have

$$v_j(x,\tau) = \sum_n v_{nj}(x) e^{-in\omega\tau}, \tag{17}$$

where $v_{nj} = v_n(x)|B_j|e^{i\phi_j \operatorname{sgn} n}$. The spectral component with $n=0$ is zero because the bulk of the solid is assumed to be at rest. Following the approach for isotropic media,[27] it can be shown that

$$\frac{\partial v_j}{\partial x} = \frac{C}{c^2T^2} \frac{\partial}{\partial \tau} \int \int_{-T/2}^{T/2} L(\phi_j, \tau-\tau', \tau-\tau'') \times v_j(x,\tau') v_j(x,\tau'')\, d\tau'\, d\tau'', \tag{18}$$

where $T = 2\pi/\omega$ and $C = -1/2\rho c^2|B_j|$. The kernel of the integral is

$$L(\phi_j, \tau', \tau'') = \sum_{l,m} P_{lm}(\phi_j) Q_{lm}(\phi_j) S_{lm} e^{-il\omega\tau'} e^{-im\omega\tau''}, \tag{19}$$

where

$$P_{lm}(\phi_j) = i(l+m)e^{i\phi_j \operatorname{sgn}(l+m)}, \tag{20a}$$

$$Q_{lm}(\phi_j) = -\operatorname{sgn}(lm) e^{-i\phi_j \operatorname{sgn} l} e^{-i\phi_j \operatorname{sgn} m}. \tag{20b}$$

As mentioned in Sec. II A, linear theory determines only the relative phases of the B_j. Without loss of generality, choose the absolute phase of one velocity component v_{j_0} (i.e., set j equal to a specific index we shall call j_0) to be a negative imaginary number so that the result corresponds to the phase convention used previously in the theory for isotropic media.[12] Under this condition, the kernel can be shown[2] to reduce to

$$L(\phi_{j_0} = -\pi/2, \tau', \tau'') = \sum_{l,m} |l+m| S_{lm} e^{-i\omega\tau'} e^{-im\omega\tau''}. \tag{21}$$

Other velocity components (for which $j \neq k$) may be obtained from integral transforms of the $v_j(x,\tau)$ component. From Eqs. (17), it follows that the nth spectral amplitudes of the jth and kth velocity components are related by

$$v_{nk}(x) = \begin{cases} (B_k/B_j) v_{nj}(x) & \text{for } n>0, \\ (B_k^*/B_j^*) v_{nj}(x) & \text{for } n<0, \end{cases} \quad (j \neq k). \tag{22}$$

Equations (22) imply that the velocity waveform components in the x_j and x_k directions are related by[15]

$$v_k(x,\tau) = \operatorname{Re}(B_k/B_j) v_j(x,\tau) - \operatorname{Im}(B_k/B_j) \mathcal{H}[v_j(x,\tau)], \tag{23}$$

where

$$\mathcal{H}[f(\tau)] = \frac{1}{\pi} Pr \int_{-\infty}^{\infty} \frac{f(\tau')}{\tau'-\tau} d\tau' \tag{24}$$

defines the Hilbert transform, and Pr indicates the Cauchy principal value of the integral.

III. NUMERICAL RESULTS

The focus of this article is the propagation of nonlinear surface waves in cubic crystals. Cubic crystals are chosen not because of any limitation of the underlying theory [Eqs. (1)–(6) apply for arbitrary anisotropy], but because they exhibit the simplest types of fully three-dimensional anisotropy, and there is considerable experimental data available for their material properties, especially higher-order elastic constants. Eight nonpiezoelectric crystals were chosen for study to provide a variety of anisotropy ratios.

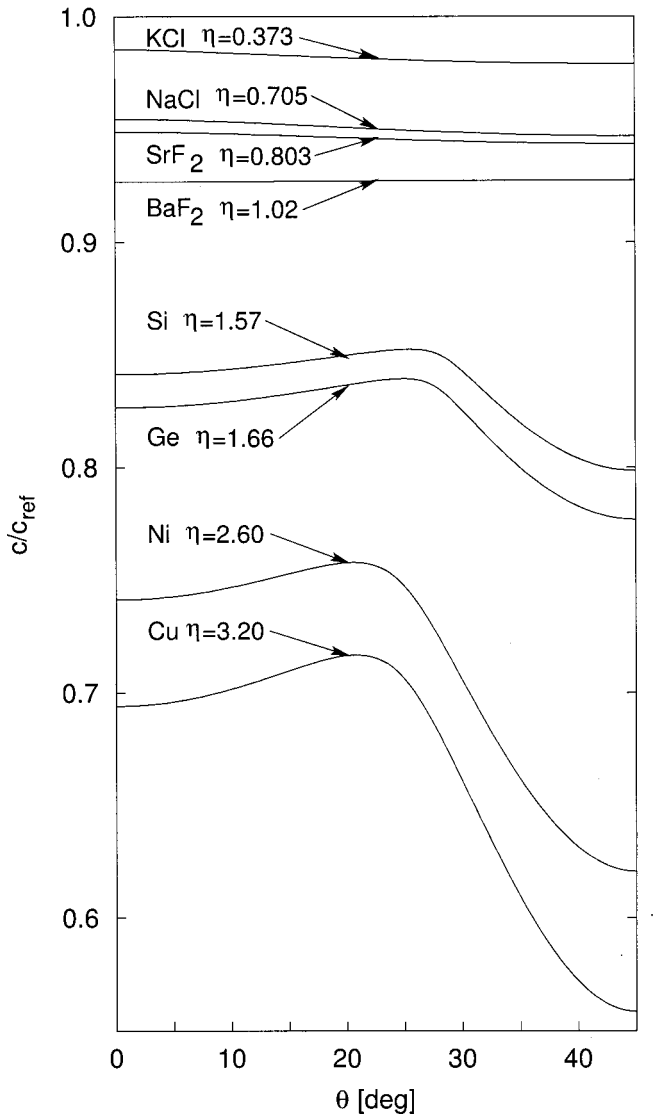


FIG. 2. Dependence of SAW speed on direction of propagation in the (001) plane for selected materials. The SAW speed of each material is measured relative to $c_{\text{ref}}=(c_{44}/\rho)^{1/2}$, and the angle θ gives the direction of the wave vector relative to $\langle 001\rangle$. Because of the symmetries of this cut, the wave speeds are symmetric about $\theta=45°$ and periodic every $\Delta\theta=90°$.

A. Linear effects

The variation of the linear wave speed as a function of angle may be characterized by the anisotropy ratio[24]

$$\eta=\frac{2c_{66}}{c_{11}-c_{12}}, \tag{25}$$

where $c_{66}=c_{44}$ for cubic symmetry. This definition is constructed such that $\eta=1$ for an isotropic material. The SAW speeds for many materials group conveniently by anisotropy ratio. A larger anisotropy ratio implies a slower SAW speed relative to the fast transverse bulk wave speed. Because of this grouping, selected materials can be shown to be characteristic of many others.[24]

Figure 2 shows the SAW speed of selected materials as a function of propagation direction. These curves were computed using density data from review articles by Pies *et al.*[28] and Eckerlin *et al.*,[29] and second-order elastic constant data from a review article by Hearmon.[30] The direction of propagation is measured by angle θ from $\langle 100\rangle$, and the wave speed for each material is scaled by $c_{\text{ref}}=(c_{44}/\rho)^{1/2}$ for that material. For the (001) surface cut, c_{ref} is also the speed of the fast transverse bulk wave, which is constant for all directions in this plane. Because the normal to the (001) plane is a fourfold symmetry axis, the SAW speed is periodic every $\Delta\theta=90°$. In addition, the $\langle 110\rangle$ direction is a twofold symmetry axis, and therefore the SAW speed is symmetric about that direction ($\theta=45°$ in Fig. 2). In most cases (and for all cases shown here) the speeds group by anisotropy ratio, with materials possessing lower anisotropy ratios having higher relative SAW speeds. Materials with $\eta\approx 1$ are nearly isotropic and hence have nearly constant SAW speed for all directions.

For all materials, the direction $\theta=0°$ is a pure mode, and the particle trajectories are confined to the sagittal plane. Some materials have additional pure modes in other directions, but in these cases the SAW often has a transverse component (e.g., Ni, Cu, Si, Ge). In some materials, the SAW speed approaches the transverse bulk wave speed as the direction approaches $\theta=45°$.[24] This effect is the source of the dip in the SAW speeds of Ni, Cu, Si, and Ge observed in Fig. 2. The SAW becomes a shear horizontal bulk wave at $\theta=45°$, and the degeneracy gives rise to a pseudosurface wave mode in the same direction. Because the nonlinear theory described above does not apply to pseudosurface waves, modes propagating in these directions are not considered further.

All of the aforementioned effects pertain to linear SAWs; they are mentioned here primarily to provide a context for the nonlinear effects discussed next.

B. Nonlinear effects

It is convenient to introduce the dimensionless quantities

$$V_n=v_n/v_0, \quad X=x/x_0, \quad A_n=\alpha_n x_0, \tag{26}$$

where v_0 is a characteristic velocity of the SAW and x_0 is given by Eq. (14). Equation (3) becomes

$$\frac{dV_n}{dX}+A_nV_n=\frac{n^2}{8|S_{11}|}\left(\sum_{m=1}^{n-1}S_{m,n-m}V_mV_{n-m}-2\sum_{m=n+1}^{N}S^*_{n,m-n}V_mV^*_{m-n}\right), \tag{27}$$

where N is the number of harmonics retained in the calculation. The boundary conditions corresponding to Eq. (7) are

$$V_1=1, \quad V_{n>1}=0. \tag{28}$$

It also proves useful to define the dimensionless nonlinearity matrix

$$\hat{S}_{lm}=-S_{lm}/c_{44}. \tag{29}$$

The negative sign is introduced to be consistent in sign with the nonlinearity matrix elements R_{lm} used for nonlinear Rayleigh waves.[31] In all cases, the figures throughout this article report matrix elements defined by Eq. (29). Because the

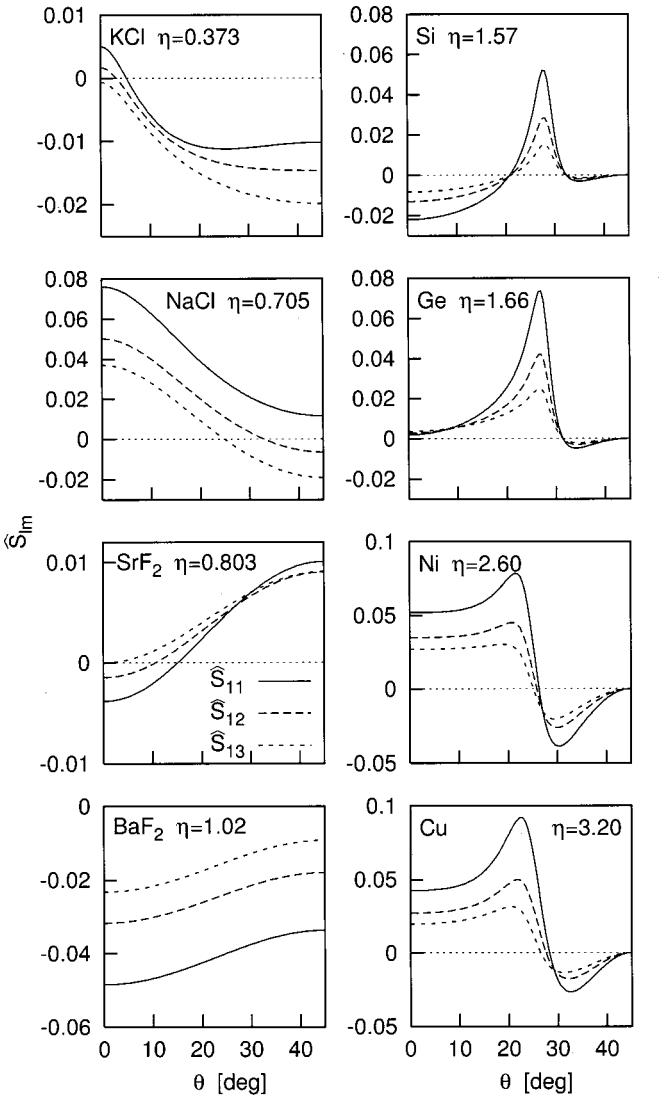


FIG. 3. Dependence of nonlinearity matrix elements on direction of propagation in the (001) plane in selected materials. The solid, long-dashed, and short-dashed lines correspond to $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$, respectively.

(001) surface is a plane of mirror symmetry, the nonlinearity matrix coefficients are all real-valued.

1. General study

Figure 3 shows plots of the nonlinearity matrix elements for a variety of crystals. Each graph plots $\hat{S}_{11}$ (solid), $\hat{S}_{12}$ (long-dashed), and $\hat{S}_{13}$ (short-dashed) over the range of directions $\theta=0°$ to $\theta=45°$. While plotting only three elements of the nonlinearity matrix certainly does not provide a full description of the nonlinear properties of the SAW, it can give a good idea of the evolution in many cases, as shown later. Moreover, the leading order terms in perturbation solutions for second, third, and fourth harmonic generation are proportional to $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$, respectively.

Four groups of crystals are shown in Fig. 3, with two different materials of similar structure in each group. The materials are ordered by increasing anisotropy ratio from top to bottom within each column (recall Fig. 2). The densities and second-order elastic constants used to construct Fig. 3 are the same as in Fig. 2, and the third-order elastic constants are listed in Table I.[32–36]

The first column of Fig. 3 shows the nonlinearity matrix elements for materials with η less than or approximately equal to unity: KCl, NaCl, SrF_2, BaF_2 ($\eta=0.373$, 0.705, 0.803, 1.02, respectively). The second column shows the nonlinearity matrix elements for materials with η greater than unity: Si, Ge, Ni, Cu ($\eta=1.57$, 1.66, 2.60, 3.20, respectively). Note that at $\theta=45°$, where the SAW and shear modes converge, $|\hat{S}_{lm}|\to 0$ in these materials. Si and KCl are chosen for investigation in detail in Secs. III B 2 and III B 3, respectively, because they provide examples of the different kinds of waveform distortion exhibited by nonlinear SAWs in the (001) plane.

2. Detailed study of silicon

Figure 4 (expanded from Fig. 3) shows the plot of $\hat{S}_{lm}$ for Si divided into three regions. The small circles on the curves correspond to the angles $\theta=0°$, 26°, and 35° discussed in detail below. In region I ($0°\leqslant\theta\leqslant 20.8°$) the nonlinearity is negative ($\hat{S}_{lm}<0$). As is shown below, this means that positive segments of the longitudinal particle velocity waveform steepen backward in space, and negative segments steepen forward (i.e., opposite to what a sound wave does in a fluid). In region II ($20.8°<\theta<32.3°$) the nonlinearity is positive ($\hat{S}_{lm}>0$), with waveform distortion the reverse of the first region. In region III ($32.3°<\theta\leqslant 45°$) the nonlinearity is again negative, although relatively weak. Observe that the linear wave speed varies by approximately 2% (see Fig. 2) over the angular range shown, whereas the changes in the nonlinearity matrix elements are of order unity. Hence not only do the nonlinear matrix elements change sign (with distinctly different waveform evolution as a result), but they also vary more widely in magnitude.

TABLE I. Third-order elastic (TOE) constants for selected nonpiezoelectric crystals. The constants are given relative to the reference frame defined by the crystalline axes in Voigt's notation with units of GPa.

Material	d_{111}	d_{112}	d_{123}	d_{144}	d_{155}	d_{156}	Source
KCl	−726	−24	+11	+23	−26	+16	Drabble *et al.* (Ref. 32)
NaCl	−843	−50	+46	+29	−60	+26	Drabble *et al.* (Ref. 32)
SrF_2	−821	−309	−181	−95.1	−175	−42.1	Alterovitz *et al.* (Ref. 33)
BaF_2	−584	−299	−206	−121	−88.9	−27.1	Gerlich (Ref. 34)
Si	−825	−451	−64	+12	−310	−64	McSkimin *et al.* (Ref. 35)
Ge	−710	−389	−18	−23	−292	−53	McSkimin *et al.* (Ref. 35)
Ni	−2032	−1043	−220	−138	−910	+70	Salama *et al.* (Ref. 36)
Cu	−1390	−778	−181	−140	−648	−16	Salama *et al.* (Ref. 36)

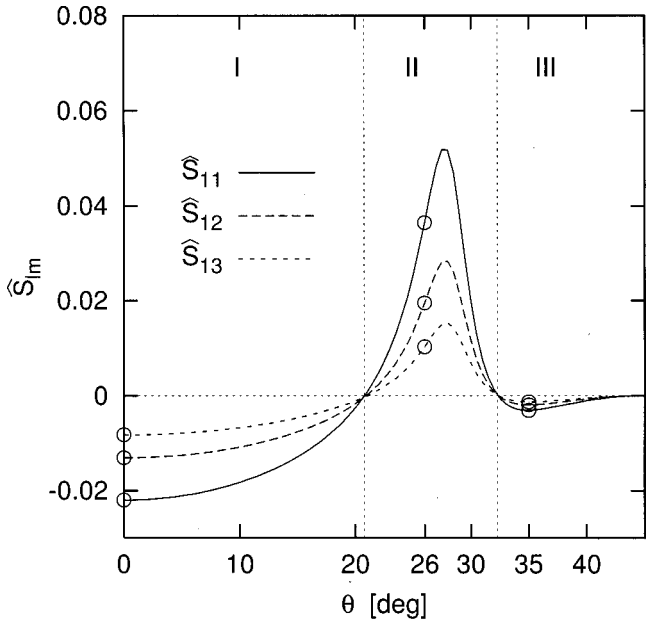


FIG. 4. Nonlinearity matrix elements $\hat{S}_{11}$, $\hat{S}_{12}$, $\hat{S}_{13}$ for Si in the (001) plane as a function of direction. Because of the symmetries of this cut, the matrix elements are symmetric about $\theta=45°$ and periodic every $\Delta\theta=90°$. The circled directions are discussed in detail in the text.

Observe that the weakening in the third region is coincident with the gradual convergence of the SAW mode and transverse bulk mode into a shear horizontal bulk wave, as seen from the wave speed plot in Fig. 2. Additional calculations[2] indicate that an increasing amount of energy moves away from the surface and into the bulk of the solid as the $\theta=45°$ direction is approached. As a result, it becomes increasingly difficult to create the surface amplitudes necessary to observe nonlinear effects in this SAW mode. (There does, however, exist a pseudosurface wave mode at a higher wave speed,[24] and it is this mode which is often excited experimentally in this region.)

Finally, Fig. 4 indicates that $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$ pass through zero near $\theta\simeq 20.8°$ and $\theta\simeq 32.3°$. Additional calculations show that while all matrix elements do not go through zero at the same angle, all significant elements are close to zero at these angles. Hence propagation is expected to be nearly linear in both these directions even for finite-amplitude waves. Because harmonic generation is suppressed, shocks do not form, or they form only over very large distances.

To investigate the effects of the sign and magnitude of $\hat{S}_{lm}$ on the velocity waveforms, simulations were performed by solving the system given by Eqs. (27) for the monofrequency source condition given in Eq. (28). The equations were integrated numerically using a fourth-order Runge-Kutta routine with $N=200$ harmonics (i.e., $1<n<N$, with $V_{-n}=V_n^*$). The nondimensional distance $X=x/x_0=1$ corresponds to approximately one shock formation distance. The absorption coefficients were chosen by assuming classical absorption due to viscosity and heat conduction, for which the quadratic frequency dependence $A_n=n^2A_1$ is obtained. The absorption for the fundamental was selected to be $A_1=\alpha_1 x_0=0.025$ to make the absorption length much larger than the shock formation distance. The effect of the absorption is then weak in comparison to that of the nonlinearity, and its primary influence is on the rise time of the shock. Once the spectra were generated, the waveforms were reconstructed using Eq. (1).

Figure 5 displays the simulation results for one direction in each region of Fig. 3. Each row shows waveform evolution in the direction specified by the angle listed in that row. The columns from left to right give the longitudinal, transverse, and vertical components of the velocity, respectively. In each direction, the waveforms are normalized such that $|V_x|^2+|V_y|^2+|V_z|^2=1$ at $X=0$, and hence the magnitudes for different directions should not be compared. Each graph of the velocity components contains waveforms at $X=0$ (short-dashed), $X=1$ (long-dashed), and $X=2$ (solid) in the retarded time frame, i.e., a frame moving at the linear SAW speed.

$\theta=0°$: This direction is in region I of Fig. 4, where the nonlinearity elements $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$ are negative. As mentioned previously, the longitudinal velocity waveform exhibits distortion with the peak receding and the trough advancing, opposite to that of a sound wave in a fluid. In addition, the longitudinal waveform exhibits the cusping near the shock front that is characteristic of SAWs. The vertical velocity waveform also exhibits the cusped peak seen in Rayleigh waves. In fact, SAWs in this direction are considered to be ''Rayleigh-type'' waves, as defined by Farnell,[24] for the following three reasons. First, due to the symmetries in this direction, $B_2=0$, and the motion is thus confined to the sagittal plane. Second, this is a pure mode direction. (Experimentally, pure mode directions are often preferred over other directions because it is typically easier to make measurements when the power flow is in the same direction as the wave vector.) Third, the principal axis of the surface particle trajectory is perpendicular to the free surface due to the 90° phase difference between B_1 and B_3. Thus, except for the fact that the amplitudes of the particle velocities do not decay purely exponentially into the solid, propagation in this direction is quite similar to the propagation of nonlinear Rayleigh waves in isotropic materials with negative nonlinearity coefficients (e.g., fused quartz[22]).

$\theta=26°$: This direction is in region II of Fig. 4, and it is the only direction in this region where there is a pure mode. $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$ are positive, and the longitudinal velocity waveform exhibits distortion with the peak of the wave advancing and the trough receding, opposite that in region I. Accordingly, the vertical velocity forms a cusped peak in the positive direction. Note that the horizontal scale has been shifted over by π rad in these waveforms, as compared to the $\theta=0°$ case.

$\theta=35°$: Here, and throughout region III of Fig. 4, $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$ are again negative, and the waveform distorts as in the $\theta=0°$ case. (Note that the horizontal scale has been shifted over by $-\pi$ rad, as compared to the $\theta=26°$ case, so as to be the same as in the $\theta=0°$ case.) This direction is not a pure mode, but it is the direction where the nonlinearity has the largest magnitude in this region. Even so, the magnitude

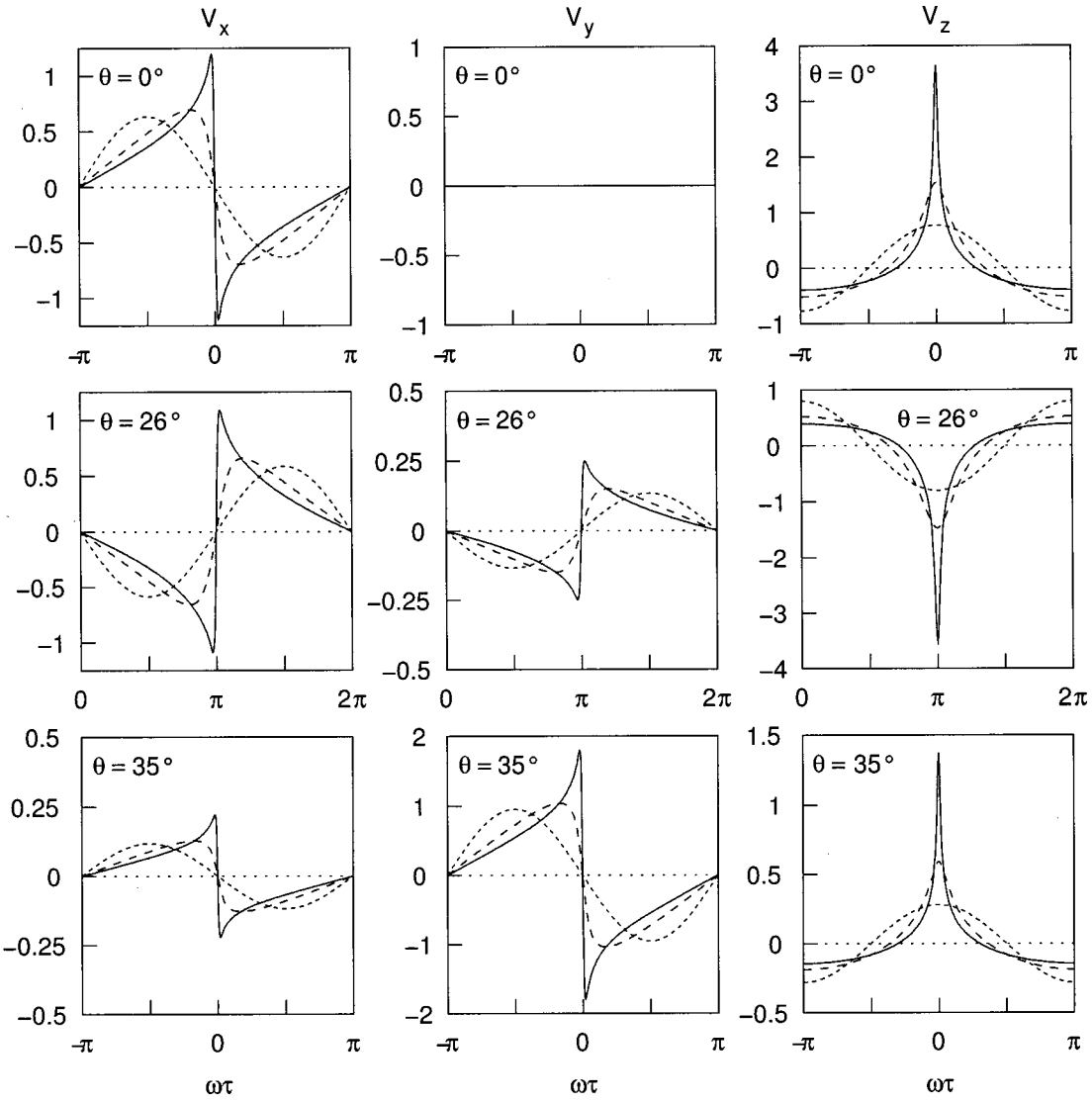


FIG. 5. Velocity waveforms in selected directions of propagation in the (001) plane of Si. The velocity components are normalized such that the initial amplitude satisfies $|V_x|^2+|V_y|^2+|V_z|^2=1$ in each propagation direction. The short-dashed, long-dashed, and solid lines correspond to propagation at distances $X=0$, $X=1$, and $X=2$, respectively.

of the nonlinearity is significantly weaker here than at $\theta=0°$, with $S_{11}(35°)/S_{11}(0°)\simeq 0.14$.

The simulations for Si presented above demonstrate both that the SAW nonlinearity varies significantly in magnitude and direction throughout the (001) cut, and that the nonlinearity matrix elements provide a map which can characterize the nature of the waveform distortion. Additional types of waveform distortion not exhibited in Si are shown in Sec. III B 3.

3. Detailed study of potassium chloride

The nonlinearity matrix elements $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$ for KCl are shown in Fig. 6 over the range $0°<\theta<10°$ (expanded in part from Fig. 3). The small circles on the curves correspond to the angles $\theta=0°$, 3.2°, and 10°, which are discussed in detail later in this work. As compared to the nonlinearity matrix elements for Si, the matrix elements for KCl are quantitatively and qualitatively different. At $\theta=0°$, $\hat{S}_{11}$ and $\hat{S}_{12}$ are positive while $\hat{S}_{13}$ is negative. As the angle from $\langle 100\rangle$ increases, $\hat{S}_{12}$ passes through zero around $\theta=3.2°$, and then finally $\hat{S}_{11}$ passes through zero around $\theta=5.2°$. The propagation at $\theta=5.2°$ is similar to that for $\theta=20.8°$ and $\theta=32.3°$ in Si, i.e., little harmonic generation occurs. At larger angles, the matrix elements $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$ are negative although, unlike Si, here the magnitude of $\hat{S}_{11}$ is less than the magnitudes of $\hat{S}_{12}$ and $\hat{S}_{13}$. Also unlike Si, the SAW modes do not converge with the transverse bulk mode at $\theta=45°$. Finally, here the percentage variation in the nonlinearity matrix elements as a function of angle again exceeds the corresponding variations in wave speed (compare with Fig. 2).

Figure 7 shows three additional types of velocity waveform distortion. Figure 7 has the same format as Fig. 5 except that the third and fourth columns show, respectively, the spectra at the locations corresponding to the waveforms, and the harmonic propagation curves for the first five harmonics. Also, the V_y waveforms are omitted because they are either zero or similar to the V_x waveforms.

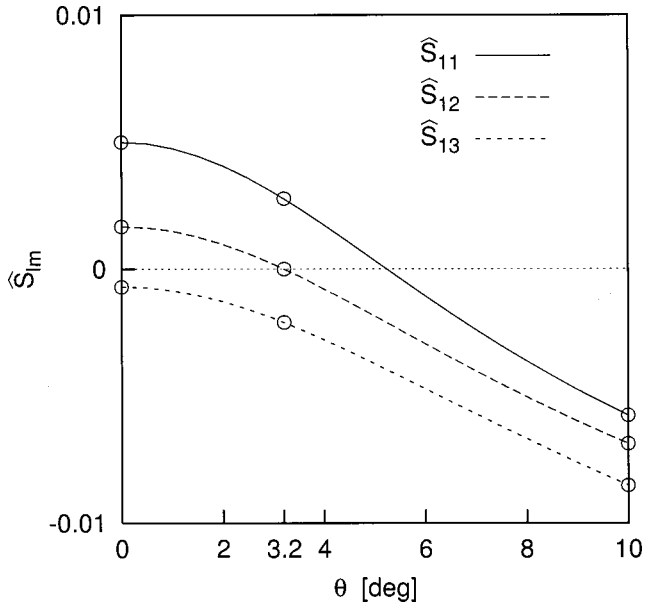


FIG. 6. Nonlinearity matrix elements $\hat{S}_{11}$, $\hat{S}_{12}$, $\hat{S}_{13}$ for KCl in the (001) plane as a function of direction. Because of the symmetries of this cut, the matrix elements are symmetric about $\theta=45°$ and periodic every 90°. The circled directions are discussed in detail in the text.

$\theta=0°$: In this direction, $\hat{S}_{11}$ and $\hat{S}_{12}$ are positive, and $\hat{S}_{13}\simeq-7.0\times10^{-4}$ is negative but close to zero and smaller than $\hat{S}_{11}$ and $\hat{S}_{12}$ in magnitude ($|\hat{S}_{13}|/|\hat{S}_{11}|\simeq0.14$, $|\hat{S}_{13}|/|\hat{S}_{12}|\simeq0.42$). The matrix element $\hat{S}_{13}$ indicates the coupling strength between the fundamental and third harmonic to generate the fourth harmonic. Energy that is transferred to the third harmonic from lower harmonics is therefore not as easily transferred to the fourth harmonic. The third harmonic (short-dashed) curve exceeds the second harmonic (long-dashed) in amplitude around $X=3$, while the fourth harmonic grows more slowly initially. Another consequence of a small value of $\hat{S}_{13}$ is that shock formation does not occur. For example, while the longitudinal waveform is distorting in a "positive" way, with peaks advancing and troughs receding, it has not yet formed a shock. Moreover, as seen in the frequency spectrum, relatively little energy is transferred to the higher spectral components as the wave propagates (contrast with the spectrum in the $\theta=10°$ direction). Additional calculations indicate that at distances $X>2$ the longitudinal velocity waveform steepens more but never forms a shock. Hence this is another direction at which shock suppression occurs. In contrast with Si, however, the harmonic suppression is less severe and occurs for a harmonic higher than the second.

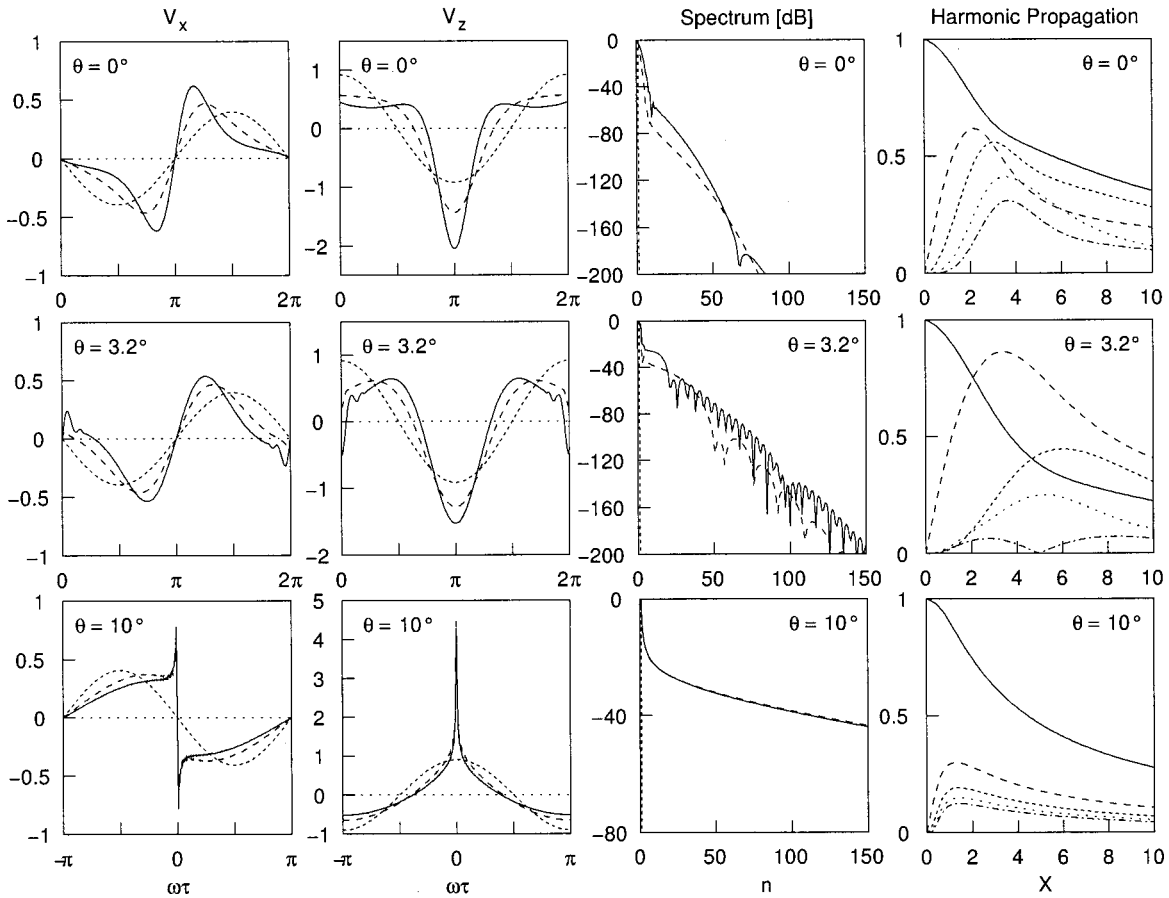


FIG. 7. Velocity waveforms in selected directions of propagation in the (001) plane of KCl. The velocity components are normalized such that initial amplitude satisfies $|V_x|^2+|V_y|^2+|V_z|^2=1$ in each propagation direction. In the velocity and spectrum plots, the short-dashed, long-dashed, and solid lines correspond to propagation at distances $X=0$, $X=1$, and $X=2$, respectively. The harmonic propagation curves plot the spectral amplitudes $|V_1|$ (solid), $|V_2|$ (long-dashed), $|V_3|$, (short-dashed), $|V_4|$ (dotted), and $|V_5|$ (dot-dashed) as a function of distance.

Finally, this is a pure mode direction, and it may be amenable to measurement.

$\theta=3.2°$: Here $\hat{S}_{11}$ is positive, $\hat{S}_{12}\simeq -4\times10^{-6}$ is close to zero, and $\hat{S}_{13}$ is negative. In addition $\hat{S}_{12}$ is much smaller than $\hat{S}_{11}$ in magnitude ($|\hat{S}_{12}|/|\hat{S}_{11}|\simeq1.5\times10^{-3}$). The matrix element $\hat{S}_{12}$ indicates the coupling strength between the fundamental and second harmonic associated with generation of the third harmonic. The result of the small value of $\hat{S}_{12}$ is highly efficient energy transfer from the fundamental to the second harmonic (because the transfer from the second to the third harmonic is impeded), which in turn causes rapidly depleting energy from the fundamental (see the Appendix of Ref. 1 for additional discussion). The inefficient transfer of energy out of the lowest harmonics leads to other unusual phenomena, including the amplitude of the third harmonic (short-dashed) exceeding the fundamental past $X=5$, and the suppression of fourth harmonic (dotted) around the same location. The complicated interaction of the lowest harmonics is also reflected in the spectra, which exhibit many maxima and minima and also show that relatively little energy is transferred to the higher spectral components. The waveforms do not exhibit shock formation but do show some higher frequency oscillations due to the atypical energy transfer. Additional calculations show that at distances $X>2$ the waveform does not form a shock, and the higher frequency oscillations grow in extent and magnitude. In summary, the simulations identify the $\theta=3.2°$ direction as another direction of shock suppression, although with a still different character than both of those described for Si or for the $\theta=0°$ case in KCl.

$\theta=10°$: Here $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$ are negative but, unlike for $\theta=0°$ and $\theta=35°$ in Si, here the magnitude $|\hat{S}_{11}|$ is less than neighboring elements (e.g., $|\hat{S}_{11}|/|\hat{S}_{12}|\simeq0.84$ and $|\hat{S}_{11}|/|\hat{S}_{12}|\simeq0.68$). This inversion in the magnitudes of the matrix elements causes energy to be transferred to the higher harmonics more efficiently. In turn, this increased rate of energy transfer results in significantly sharper cusping in the waveforms. The harmonic propagation curves also show a steeper decline in the amplitude of the fundamental and a steeper increase in the other harmonics. All directions $10°<\theta<45°$ show this same type of distortion.

Before concluding this section, comments are in order regarding the dependence of the aforementioned results on the choice of third-order elastic constant data. To evaluate the effects of using the same second-order elastic constants but different third-order elastic constants, the nonlinearity matrix elements $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$ corresponding to the data of Drabble and Strathen[32] (used in the KCl simulations above) are plotted together with those of Chang[37] in Fig. 8. Both sets show the same general trend, but the locations of the zero crossings occur for slightly smaller values of θ in the Chang data. Because Hamilton *et al.*[1] used the Chang data to perform their simulations for KCl, their paper shows the kind of waveform distortion seen for $\theta=3.2°$ in this article at $\theta=0°$. An examination of Chang's paper shows that only three of the six TOE constants were determined experimentally (d_{144}, d_{155}, d_{456}) while the other elastic constants

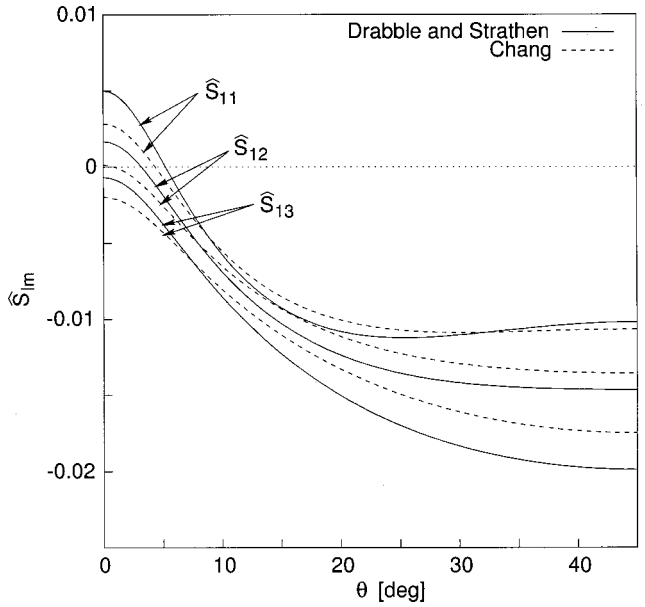


FIG. 8. Comparison of selected nonlinearity matrix elements calculated from third-order elastic constant data of Drabble and Strathen (Ref. 32) and Chang (Ref. 37) for propagation in the (001) plane of KCl. Second-order elastic constant data are taken from Hearmon (Ref. 30) in both cases.

(d_{111}, d_{122}, d_{123}) were computed by assuming the Cauchy relations $d_{123}=d_{456}=d_{144}$ and $d_{112}=d_{166}$. In contrast, Drabble and Strathen measured all six constants and showed that not all the Cauchy relations hold (see Table I). (Although not shown here, a similar comparison[2] was performed for the Si case described in Sec. III B 2. The general trends remain the same although the exact amplitude and zero crossings change by a small amount.) The lesson here is that the nonlinearity matrix elements, and therefore the waveform distortion, can be rather sensitive to changes in the third-order elastic constants. Therefore care should be exercised in making detailed predictions at specific directions without accurate third-order elastic constants.

IV. SUMMARY

This article has investigated the properties of nonlinear SAWs in the (001) plane of selected crystals. Si and KCl were chosen for detailed study. It is found that the nature of the nonlinearity is often very sensitive to changes in direction, and the relative variations can be larger than those occurring in the linear wave speed. For most cases, plotting the first few nonlinearity matrix elements as a function of direction can provide a guide to the nature of the nonlinear effects. While particular directions of high symmetry exhibit harmonic generation and waveform distortion similar to Rayleigh waves, several other effects have been identified. These include the existence of regions of positive and negative nonlinearity within the same cut, several varieties of shock suppression where the generation of particular harmonics is suppressed and shock formation does not occur, and directions in which harmonic generation is increased and energy is more efficiently transferred from the fundamental to higher harmonics. As a result of the last two effects, the simple estimate of the shock formation given by Eq. (14) may not be valid or accurate in some regions. The choice of different

experimental third-order elastic constants as input to the simulations is shown to affect the detailed predictions of the nonlinearity matrix elements in any given direction, but not the general trends over the whole angular range.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. Discussions with Yu. A. Ilinskii and E. A. Zabolotskaya are also gratefully acknowledged.

[1] M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "Nonlinear surface acoustic waves in crystals," J. Acoust. Soc. Am. **105**, 639–651 (1999).
[2] R. E. Kumon, "Nonlinear surface acoustic waves in cubic crystals," Ph.D. dissertation, The University of Texas at Austin, 1999.
[3] R. E. Kumon, M. F. Hamilton, P. Hess, A. Lomonosov, and V. G. Mikhalevich, "Dependence of surface acoustic wave nonlinearity on propagation direction in crystalline silicon," in *Nonlinear Acoustics at the Turn of the Millennium*: *Proceedings of the 15th International Symposium on Nonlinear Acoustics*, Vol. 524 of AIP Conference Proceedings, edited by W. Lauterborn and T. Kurz (American Institute of Physics, Melville, NY, 2000), pp. 265–268.
[4] D. F. Parker, "Nonlinear surface acoustic waves and waves on stratified media," in *Nonlinear Waves in Solids*, No. 341 in *International Centre for Mechanical Sciences Courses and Lectures*, edited by A. Jeffrey and J. Engelbrecht (Springer-Verlag, New York, 1994), pp. 289–347.
[5] A. P. Mayer, "Surface acoustic waves in nonlinear elastic media," Phys. Rep. **256**, 237–366 (1995).
[6] M. Planat, "Multiple scale analysis of the nonlinear surface acoustic wave propagation in anisotropic crystals," J. Appl. Phys. **57**, 4911–4915 (1985).
[7] R. W. Lardner, "Nonlinear surface acoustic waves on an elastic solid of general anisotropy," J. Elast. **16**, 63–73 (1986).
[8] R. W. Lardner and G. E. Tupholme, "Nonlinear surface waves on cubic materials," J. Elast. **16**, 251–265 (1986).
[9] D. F. Parker, "Waveform evolution for nonlinear surface acoustic waves," Int. J. Eng. Sci. **26**, 59–75 (1988).
[10] D. F. Parker and E. A. David, "Nonlinear piezoelectric surface waves," Int. J. Eng. Sci. **27**, 565–581 (1989).
[11] M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "Nonlinear surface wave propagation in crystals," in *Nonlinear Acoustics in Perspective: Proceedings of the 14th International Symposium on Nonlinear Acoustics*, edited by R. J. Wei (Nanjing U.P., Nanjing, China, 1996), pp. 64–69.
[12] E. A. Zabolotskaya, "Nonlinear propagation of plane and circular Rayleigh waves," J. Acoust. Soc. Am. **91**, 2569–2575 (1992).
[13] E. Yu. Knight, M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "General theory for the spectral evolution of nonlinear Rayleigh waves," J. Acoust. Soc. Am. **102**, 1402–1417 (1997).
[14] M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "Nonlinear surface acoustic waves," in *Nonlinear Acoustics at the Turn of the Millennium*: *Proceedings of the 15th International Symposium on Nonlinear Acoustics*, Vol. 524 of AIP Conference Proceedings, edited by W. Lauterborn and T. Kurz (American Institute of Physics, Melville, NY, 2000), pp. 55–64.
[15] V. E. Gusev, W. Lauriks, and J. Thoen, "New evolution equations for the nonlinear surface acoustic waves on an elastic solid of general anisotropy," J. Acoust. Soc. Am. **103**, 3203–3215 (1998).
[16] G. D. Meegan, M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "Nonlinear Stoneley and Scholte waves," J. Acoust. Soc. Am. **106**, 1712–1723 (1999).
[17] A. Lomonosov and P. Hess, "Laser excitation and propagation of nonlinear surface acoustic wave pulses," in *Nonlinear Acoustics in Perspective: Proceedings of the 14th International Symposium on Nonlinear Acoustics*, edited by R. J. Wei (Nanjing U.P., Nanjing, China, 1996), pp. 106–111.
[18] Al. A. Kolomenskii, A. M. Lomonosov, R. Kuschnereit, P. Hess, and V. E. Gusev, "Laser generation and detection of strongly nonlinear elastic surface pulses," Phys. Rev. Lett. **79**, 1325–1328 (1997).
[19] A. A. Kolomenskii and H. A. Schuessler, "Characterization of isotropic solids with nonlinear surface acoustic wave pulses," Phys. Rev. B **63**, 085413 (2001).
[20] A. A. Kolomenskii and H. A. Schuessler, "Nonlinear compression of giant surface acoustic wave pulses," Phys. Lett. A **280**, 157–161 (2001).
[21] A. Lomonosov and P. Hess, "Effects of nonlinear elastic surface pulses in anisotropic silicon crystals," Phys. Rev. Lett. **83**, 3876–3879 (1999).
[22] A. Lomonosov, V. G. Mikhalevich, P. Hess, E. Yu. Knight, M. F. Hamilton, and E. A. Zabolotskaya, "Laser-generated nonlinear Rayleigh waves with shocks," J. Acoust. Soc. Am. **105**, 2093–2096 (1999).
[23] R. E. Kumon, M. F. Hamilton, Yu. A. Il'inskii, E. A. Zabolotskaya, P. Hess, A. Lomonosov, and V. G. Mikhalevich, "Pulsed nonlinear surface acoustic waves in crystals," in *Proceedings of the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America*, edited by P. K. Kuhl and L. A. Crum (Acoustical Society of America, Woodbury, NY, 1998), Vol. 3, pp. 1557–1558.
[24] G. W. Farnell, "Properties of elastic surface waves," in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1970), Vol. 6, pp. 109–166.
[25] M. F. Hamilton and D. T. Blackstock, eds., *Nonlinear Acoustics* (Academic, New York, 1998).
[26] E. Yu. Knight, M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "On Rayleigh wave nonlinearity, and analytical approximation of the shock formation distance," J. Acoust. Soc. Am. **102**, 2529–2535 (1997).
[27] M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "Local and nonlocal nonlinearity in Rayleigh waves," J. Acoust. Soc. Am. **97**, 882–890 (1995).
[28] W. Pies and A. Weiss, in *Crystal Structure Data of Inorganic Compounds*, *Part a: Key Elements: F, Cl, Br, I*, Vol. III/7a of *Landolt-Börnstein*, New Series, edited by K.-H. Hellwege and A. M. Hellwege (Springer-Verlag, New York, 1973).
[29] P. Eckerlin and H. Kandler, in *Structure Data of Elements and Intermetallic Phases*, Vol. III/6 of *Landolt-Börnstein, New Series*, edited by K.-H. Hellwege and A. M. Hellwege (Springer-Verlag, New York, 1971).
[30] R. F. S. Hearmon, "The elastic constants of crystals and other anisotropic materials," in *Elastic, Piezoelectric, Pyroelectric, Piezooptic, Electroptic Constants, and Nonlinear Dielectric Susceptibilities of Crystals*, Vol. III/11 of *Landolt-Börnstein, New Series*, edited by K.-H. Hellwege and A. M. Hellwege (Springer-Verlag, New York, 1979), pp. 1–244.
[31] D. J. Shull, M. F. Hamilton, Yu. A. Il'insky, and E. A. Zabolotskaya, "Harmonic generation in plane and cylindrical nonlinear Rayleigh waves," J. Acoust. Soc. Am. **94**, 418–427 (1993).
[32] J. R. Drabble and R. E. B. Strathen, "The third-order elastic constants of potassium chloride, sodium chloride and lithium fluoride," Proc. Phys. Soc. London **92**, 1090–1095 (1967).
[33] S. Alterovitz and D. Gerlich, "Third-order elastic moduli of strontium fluoride," Phys. Rev. B **1**, 2718–2723 (1970).
[34] D. Gerlich, "Third-order elastic moduli of barium fluoride," Phys. Rev. **168**, 947–952 (1968).
[35] H. J. McSkimin and P. Andreatch, Jr., "Measurement of third-order moduli of silicon and germanium," J. Appl. Phys. **35**, 3312–3319 (1964).
[36] K. Salama and G. A. Alers, "The composition dependence of the third-order elastic constants of the Cu-Ni system," Phys. Status Solidi A **41**, 241–247 (1977).
[37] Z.-P. Chang, "Third-order elastic constants of NaCl and KCl crystals," Phys. Rev. **140**, 1788–1799 (1965).

Energy: Converting from acoustic to biological resource units

Kelly J. Benoit-Bird[a] and Whitlow W. L. Au
Hawaii Institute of Marine Biology, P.O. Box 1106, Kailua, Hawaii 96734

(Received 26 June 2001; revised 14 January 2002; accepted 18 February 2002)

Acoustic backscattering strength is often used as an index of biomass; however, the relationship between these variables has not been directly validated. Relationships were investigated between acoustic cross section at 200 kHz, measured as part of a previous study, and measured values of length, biovolume, dry weight, ash-free dry weight, and caloric content of the same individual specimens. Animals were part of the Hawaiian mesopelagic boundary community and included shrimps, squids, and myctophid fishes. The strong relationships found between all the variables measured make it possible to approximate any one variable from the measured values of others within a class of animals. The data show that for these midwater animals, acoustic scattering can be used as an index of biomass. Dorsal-aspect acoustic cross section at 200 kHz predicted dry weight and ash-free dry weight at least as well as did body length, a standard predictor. Dorsal-aspect acoustic cross section at 200 kHz was also a strong predictor of total caloric content. The relationship between dorsal-aspect acoustic cross section and caloric content of Hawaiian mesopelagic animals was linear and additive. Consequently, it is possible to directly convert acoustic energy from these animals to organic resource units without having knowledge of the size distribution of the populations being studied.
[DOI: 10.1121/1.1470505]

PACS numbers: 43.30.Sf, 43.20.Fn [DLB]

I. INTRODUCTION

The goal of the majority of population and community surveys has been to assess biomass and its distribution. Biomass is measured in many ways: wet weight, dry weight, ash-free dry weight, biovolume, chlorophyll α for primary production, protein for secondary production, carbon, ATP, and energy content (calories). In every case, biomass is a measure of organic resources or available energy. Acousticians and fisheries biologists attempt to assess biomass acoustically, some using acoustic volume backscattering as an index of biomass and sometimes attempting to convert acoustic units into measures of biomass more palatable to the biologists who use this information.

Converting acoustic measurements into biomass estimates requires samples of the population being measured (Claytor *et al.*, 1999). To estimate biomass, standard length–weight relationships (Kemper and Raat, 1997) for the taxa involved are often applied to the acoustic estimates of abundance obtained. The biological samples taken therefore need to represent the length distribution and taxonomic composition of the population examined in the acoustic survey (Bethke *et al.*, 1994). This requires that the collecting gear be unbiased in capturing the objects in the volume of water sampled (Bethke *et al.*, 1994). Such nonselective collection is difficult, if not impossible (Parkinson *et al.*, 1994). The problems associated with converting acoustic energy to abundance estimates have been amply discussed (see MacLennan and Simmonds, 1992). These problems can only be further compounded by using average length–weight relationships that are often plagued by collection biases, limited data, and spatial and seasonal variations. When biological samples of the population surveyed hydroacoustically are not available, volume backscattering is often used as a direct index of biomass (Liao *et al.*, 1999). However, studies directly validating the assumption that backscatter is an appropriate measure of biological energy are rare.

Ideally, we would like to assess biomass through a direct conversion of acoustic energy to units of organic energy without an intermediate step. The goal of this work was to assess the relationship between acoustic energy, measured as part of a previous study, and various measures of biomass taken on the same individuals, including biovolume, dry weight, ash-free dry weight, and caloric energy content. This was an attempt to directly validate the use of acoustic backscatter as a measure of biomass. We also attempted to determine if acoustic scattering can be directly converted to biomass without knowledge of the size distribution of the population being surveyed, at least for animals from the Hawaiian mesopelagic boundary community—a near-shore scattering layer community of small fish, shrimp, and squid (Reid *et al.*, 1991). This is a biological perspective on assessing biomass for this noncommercial animal community rather than a more traditional fisheries acoustics perspective. Ultimately, the units obtained are the same as those estimated in many fisheries acoustics studies; however, our method of converting between acoustic measures and biomass is novel and more direct.

II. METHODS

Trawling for micronektonic animals was conducted using a 2-m opening Isaacs–Kidd midwater trawl (IKMT) during two cruises in May and July of 2000 aboard the NOAA ship TOWNSEND CROMWELL. The trawl was towed obliquely

[a]Electronic mail: benoit@hawaii.edu

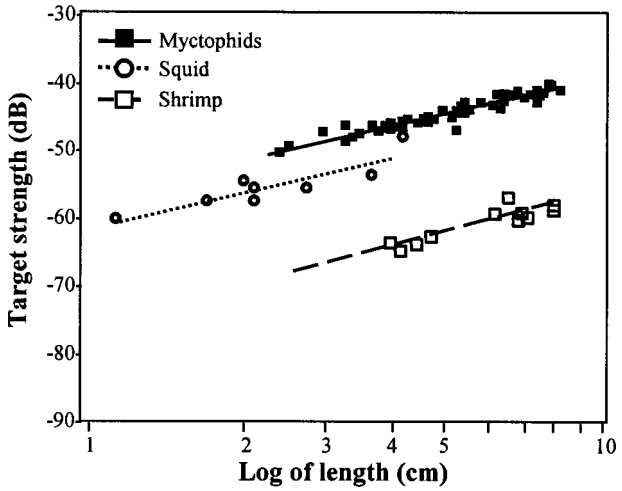


FIG. 1. The relationship between fish standard length, shrimp total length, squid mantle length, and animal target strength (redrawn from Benoit-Bird and Au, 2001).

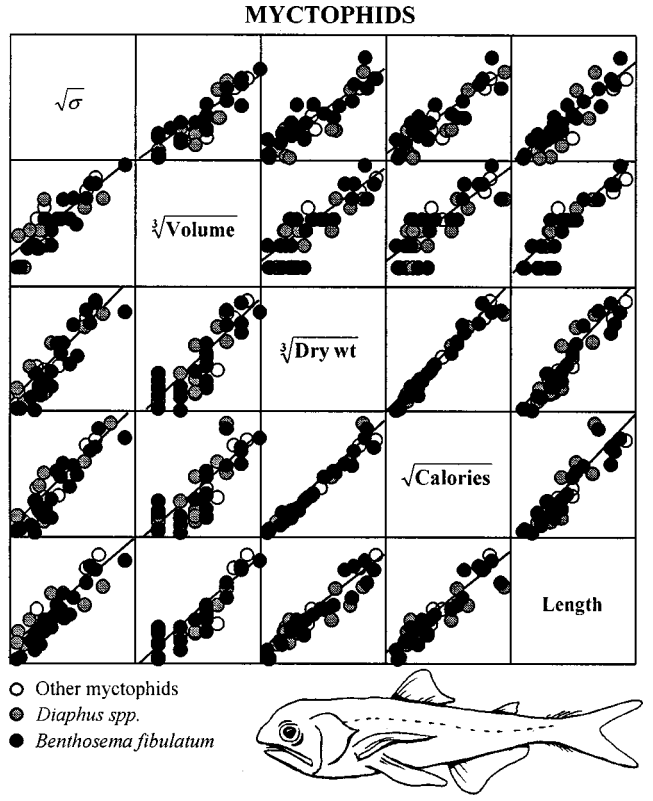


FIG. 2. Matrix of regressions for standardized measures of biomass, fish standard length, and the square root of dorsal-aspect acoustic cross section (σ) of myctophid fish. All possible combinations are shown with each variable as both the x and y axis.

for 20 to 30 min, reaching a maximum depth of 200 m. The ship was traveling between 3 and 4 knots with wire sent out at 25 m per min. The dorsal-aspect target strengths at 200 kHz, shown in Fig. 1, of the various live animals from the mesopelagic boundary community were measured as part of a previous study (Benoit-Bird and Au, 2001). The standard length of fish (the distance between the tip of the snout and the rear end of the caudal peduncle), the mantle length of the dorsal side of the squids, and the total length of the fish, squid, and shrimp species were measured with vernier calipers to the nearest 1 mm. Animals were then identified to species and frozen for later analysis.

After returning to the laboratory, the displacement volume to the nearest 0.5 ml of each individual animal was measured in a graduated cylinder of appropriate size for the individual. Animals were then homogenized in a small blender with distilled water added to facilitate even mixing. Homogenized samples were freeze-dried to remove all water without allowing the loss of volatile substances (Paine, 1971). Dry weight of the entire sample was then measured and the sample was further homogenized using a mortar and pestle.

Two 10–20-mg subsamples of each fully homogenized animal were compressed into pellet form with a Parr pellet press. Calorie values of the two subsamples were determined using a Kipp and Zonen BD40 Gentry microbomb oxygen calorimeter attached to a chart recorder, using standard methods (Paine, 1966). The samples were run in random order along with three 10–20-mg benzoic acid standards. If two subsamples disagreed by more than 3%, a third sample was run.

The percent of ash of each animal was also determined. Glass filters were heated to 500 °C for 4 h to remove any biological residues, cooled in a dessicator, and then weighed. Two freeze-dried subsamples of each animal were weighed and placed on these preweighed glass filters in aluminum foil dishes. The samples were then heated for 4 hours at 500 °C, cooled in a dessicator, and the sample and filter were weighed. The weight of the filter was then subtracted from the sample plus filter weight, and the percent of the sample that remained after ashing was calculated.

To examine the relationships between the variables measured, each was standardized into a one-dimensional variable. Acoustic cross section is closely related to the square of the mesopelagic animal's length (Benoit-Bird and Au, 2001); therefore, the square root of acoustic cross section was used for comparison between variables. Because the shapes of animals change roughly the same as length, the cube root of both volume and weight was taken. Calorie content, generally, has between a square and a cubic relationship with animal length (Golley, 1961; Slobodkin and Richman, 1961). To determine which exponent was more appropriate, the relationships of the square root and cube root of caloric content for each group against length were tested. Length predicted the square root of calories better than the cube root of calories in all three animal groups, so the square root was used for comparisons.

III. RESULTS

There was strong colinearity between standardized measures of dorsal-aspect acoustic cross section (σ), volume, dry weight, ash-free dry weight, length, and calories for myctophid fishes (Fig. 2), mesopelagic shrimps (Fig. 3), and mesopelagic squids (Fig. 4). Plots of linear, pairwise relationships of all these standardized variables, except ash-free dry weight, are found in the matrices of these three figures. For example, looking at the box labeled ''Volume,'' the graph above this box shows the cube root of volume on the x axis

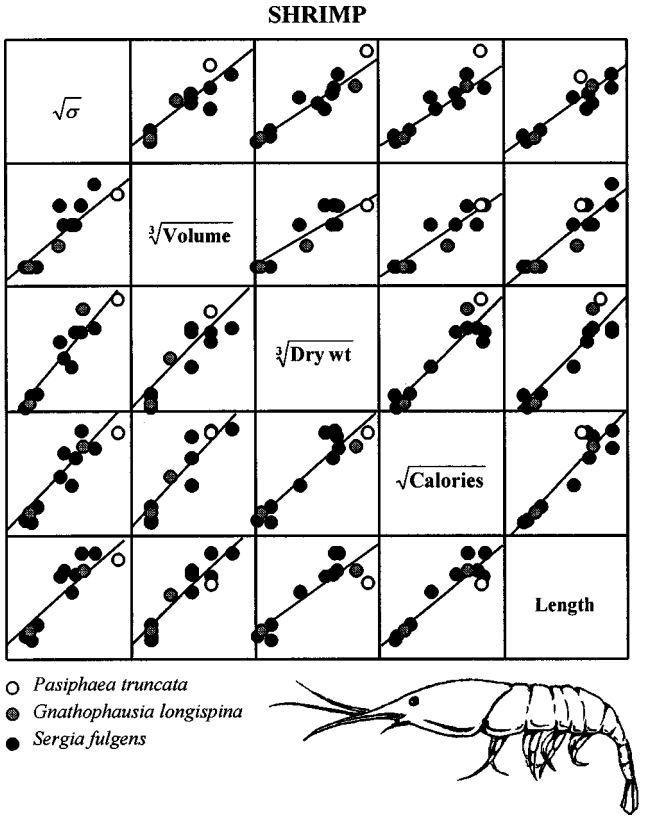


FIG. 3. Matrix of regressions for standardized measures of biomass, shrimp total length, and the square root of dorsal-aspect acoustic cross section of mesopelagic shrimps. All possible combinations are shown with each variable as both the *x* and *y* axis.

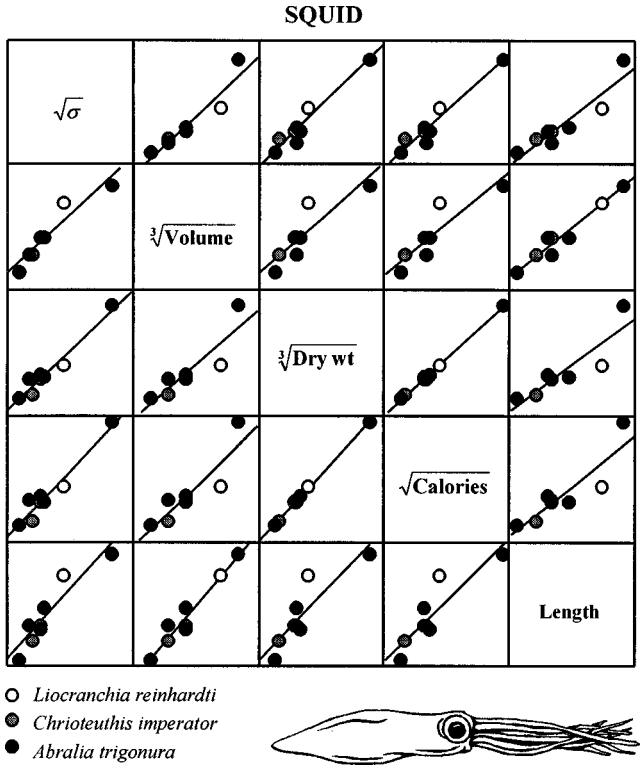


FIG. 4. Matrix of regressions for standardized measures of biomass, squid mantle length, and the square root of dorsal-aspect acoustic cross section of mesopelagic squids. All possible combinations are shown with each variable as both the *x* and *y* axis.

and the square root of σ on the *y* axis. The graph to the left of the ''Volume'' box shows the square root of σ on the *x* axis and the cube root of volume on the *y* axis. The extreme lower left graph of each matrix shows the square root of σ on the *x* axis and length on the *y* axis and so on. These figures show the linear relationship of the paired variables, not their specific values. Acoustic cross section at 200 kHz was a significant predictor of the measures of biomass taken (standardized volume, dry weight, ash-free dry weight, and calories) for all three groups and *F* tests showed that the slopes of all regressions were significant (Table I).

Total calorie content is arguably the most biologically important measure of biomass used here. The relationships between dorsal-aspect acoustic cross section and calories are shown in Figs. 5, 6, and 7. A comparison of standardized dorsal-aspect acoustic cross section with length, and standardized biomass measures (volume, dry weight, and ash-free dry weight) as independent predictors of total calorie content is shown in Table II. A blocked linear regression shows that for myctophid fish, dorsal-aspect acoustic cross section at 200 kHz is not a significantly different predictor of calorie content than body length or volume. In shrimp, both length and volume are significantly better predictors of calorie content than the dorsal-aspect acoustic cross section. For squid, dorsal-aspect acoustic cross section at 200 kHz is a significantly better predictor of calorie content than either body length or volume. For all three groups of organisms, dry weight and ash-free dry weight are nearly perfect predictors of calorie content ($r^2=0.98$, $r^2=0.97$ overall). Consequently, both are significantly better predictors of calorie content than dorsal-aspect acoustic cross section at 200 kHz in all three groups. Dorsal-aspect acoustic cross section predicts dry weight and ash-free dry weight significantly better than it predicts total calories for shrimp, but it is not significantly different in predicting both weight measures and caloric content in myctophid fishes and squids.

TABLE I. Results of regression analyses between dorsal-aspect acoustic cross section and various measures of biomass, all standardized to be one-dimensional.

Independent	Acoustic cross-section (σ)^1/2							
Dependent	Volume^1/3		Dry weight^1/3		Ash-free dry wt^1/3		Calories^1/2	
	r^2	P	r^2	P	r^2	P	r^2	P
Myctophids	0.80	<0.0001	0.83	<0.0001	0.83	<0.0001	0.82	<0.0001
Shrimp	0.67	<0.001	0.89	<0.0001	0.85	<0.0001	0.76	<0.001
Squid	0.89	<0.0001	0.89	<0.0001	0.89	<0.0001	0.89	<0.0001

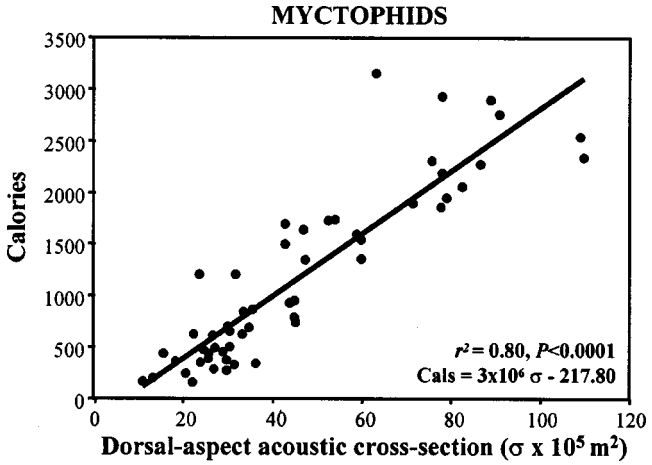


FIG. 5. Regression between dorsal-aspect acoustic cross section and total caloric content for myctophid fishes (includes all specimens).

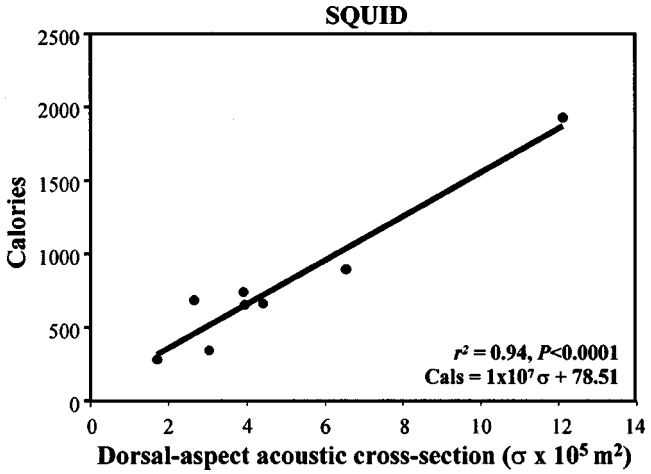


FIG. 7. Regression between dorsal-aspect acoustic cross section and total caloric content for mesopelagic squid (includes all specimens).

Energy density (calories per gram of ash-free dry weight), an important measure of food quality that allows comparison between animal groups, is summarized in Table III for the midwater animals measured. The distribution of energy density values for myctophids was skewed towards lower values. The distribution of shrimp energy density was bimodal, with modes above and below the mean energy density. The distribution of squid energy density approximated a normal distribution.

The correlation between caloric content and dorsal-aspect acoustic cross section at 200 kHz, despite large differences caloric density, suggests that energy density and acoustic scattering may be related. To examine this relationship, the energy density of animals with target strengths higher than the average for the taxonomic group to which they belong was compared with the energy density of animals with lower than average target strengths. Average was defined as the target strength predicted by the regression of length versus target strength at 200 kHz for an animal belonging to that taxonomic group (Benoit-Bird and Au, 2001). The individuals in the two groups were spread randomly throughout the size range measured. A two-tailed *t* test without assuming equal variances revealed that animals with lower than average target strength at 200 kHz had significantly higher energy density than animals with higher than average target strength at 200 kHz (Table IV).

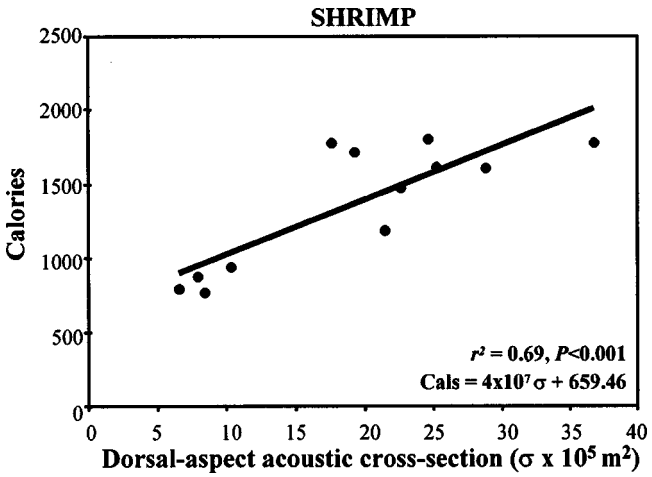


FIG. 6. Regression between dorsal-aspect acoustic cross section and total caloric content for mesopelagic shrimp (includes all specimens).

IV. DISCUSSION AND CONCLUSIONS

The strong relationship between acoustic scattering strength and biomass measures should perhaps not be surprising. Much has been written in the acoustic literature on the relationship between animal length and scattering strength (see, for example, Love, 1970). In the biological literature, animal length is routinely related to biomass measures (see, for example, Sarvala *et al.*, 1999). Both of the relationships are strong and show the same directional trend. Consequently, the relationship between acoustic scattering and biomass should also be strong.

Acoustic scattering strength was an equally good predictor of standardized measures of volume, dry weight, ash-free dry weight, and total caloric value for myctophid fish and squid. For shrimps, acoustic scattering predicted standardized dry weight and ash-free dry weight equally well, and standardized volume and calories equally well, but the weight measures were predicted significantly better than the volume or caloric content. The small sample size for both the shrimp and squid makes it difficult to interpret the meaning of this difference. Perhaps the chitinous exoskeleton of the shrimp, which made the percent of ash higher than in the other two groups, is responsible for the poorer fit of the caloric data in this group. Also, the low scattering strength of the shrimp compared with the fishes and squids means that the range of values of acoustic scattering is small compared with the range of values for fish and squid. This would make an existing relationship more difficult to detect, particularly with a small sample size. For myctophid fishes however, the predictive relationships were especially strong for caloric content, the appropriate measure of biomass for the purposes of a field study. Predictions based on dorsal-aspect acoustic backscattering, for myctophid fish and midwater squid, are at least as good as those based on length or volume, which are standard biological predictors of biomass. These data suggest

TABLE II. Results of regression analyses between one-dimensional biological and acoustic measures of individual animals and caloric content.

Independent					Dry weightˆ1/3					
Dependent	Length		Volumeˆ1/3		Caloriesˆ1/2		Ash-free dry wtˆ1/3		(σ)ˆ1/2	
	r^2	P	r^2	P	r^2	P	r^2	P	r^2	P
Myctophids	0.80	<0.0001	0.81	<0.0001	0.97	<0.0001	0.96	<0.0001	0.82	<0.0001
Shrimp	0.88	<0.0001	0.79	<0.0001	0.90	<0.0001	0.87	<0.0001	0.76	<0.001
Squid	0.84	<0.0001	0.79	<0.0001	0.99	<0.0001	0.99	<0.0001	0.89	<0.0001

that the use of acoustics for studying biomass is warranted and merits further work to validate the generality of the relationship.

Mesopelagic boundary animals from all three taxa with higher than average target strengths have lower energy densities than those with lower than average target strengths at 200 kHz. Although proximate analysis was not conducted on the animals in this study, other studies have shown a positive correlation between lipid content and caloric density (Donnelly *et al.*, 1993; Donnelly *et al.*, 1990; Ikeda, 1996; Stickney and Torres, 1989). This suggests that an increase in the proportion of lipid may be responsible for reducing the scattering strength of individuals from the boundary layer with high caloric density. This is likely because the density of marine lipids, while variable, is closer to the density of water than other body components (Donnelly *et al.*, 1990; Neighbors and Nafpaktitis, 1982; Ohshima *et al.*, 1987; Stickney and Torres, 1989), potentially making the impedance match closer, reducing the target strength. This relationship likely will not be extendable to taxa that have air-filled cavities. The presence of an air-filled cavity can increase the acoustic cross section of similarly sized animals by orders of magnitude (Medwin and Clay, 1997). The myctophid fish measured often had wax-invested or fully deflated swimbladders that did not significantly affect their backscattering strength (Benoit-Bird and Au, 2001). The shrimp and squid also lacked air-filled cavities.

This study may be the first to compare paired measures of acoustic backscatter cross section and measures of biomass. Interestingly, the strong linear relationships between one-dimensional versions of all the variables measured—length, volume, dry weight, ash-free dry weight, calories, and dorsal-aspect acoustic cross section at 200 kHz—make it possible to estimate any one of these variables from the others, within a taxonomic group. Biologists conducting biomass analyses have long reported data derived from length–weight curves, weight–calorie curves, and length–calorie curves (Cummins and Wuycheck, 1977). Acousticians doing fisheries work have applied these curves, primarily those relating length and weight, to their data to estimate biomass (Bethke *et al.*, 1994). However, the results from this study provide additional capabilities. First, these data show that in these midwater animals, acoustic scattering measurements can be used as a direct index of biomass in monospecific aggregations or when the proportion of each group within the survey area is known. Such estimation is commonly done (Liao *et al.*, 1999) but the relationship of scattering to biomass has not been directly validated. Second, the relationship between dorsal-aspect acoustic cross section and caloric content of Hawaiian mesopelagic animals is linear and additive (Foote, 1983). Consequently, in the Hawaiian mesopelagic boundary layer where the size range of animals present is very narrow, it is possible to directly convert acoustic energy to organic resource units without having knowledge of the size distribution of the population being studied. However, it is necessary to know the proportion of each biological group (myctophid fish, shrimp, and squid) because myctophid fish have higher scattering values than equivalently sized squid and shrimp (Benoit-Bird and Au, 2001), while equivalently sized animals have similar caloric values.

Our results suggest that acoustic scattering can provide a useful measure of biomass for these midwater species and can be converted to biologically relevant units without intermediate steps. However, we have no evidence that such estimates can be extrapolated to other species in other areas. The process of directly measuring energy density is very time consuming, but the measurement of dry weight is relatively simple. Individual animals need not be homogenized before drying if calorimetric analysis is not to be conducted. The only equipment necessary is a freeze-dryer (also known as a lyophilizer) and an accurate balance appropriate to the size of the sample. Biovolume is an even simpler measurement to take, requiring only a graduated cylinder and water. We suggest that measures of biomass should become a routine part of studies measuring acoustic scattering under controlled conditions. These types of paired measurements would prove useful for field studies and models of ecosystems. Such measurement of biomass has been relegated by most acousticians to biologists. However, measuring acoustic backscatter may be much more difficult for biologists than measuring biomass is for acousticians. It is crucial that acousticians participate in obtaining the important paired measurements of acoustic backscattering and biomass for other species.

TABLE III. Summary of energy density, calories per gram of ash-free dry weight, for the midwater animals measured.

	n	Minimum	Maximum	Mean	Std. deviation
Myctophids	54	3721	9071	5309	909
Shrimp	12	3823	5785	4836	569
Squid	8	4492	4949	4712	153

TABLE IV. Two-tailed t test for equality of energy density means for mesopelagic animals that have lower than average target strength and those that have higher than average target strength.

T	df	P (2-tailed)	Mean difference	Std. error difference
2.0	70.2	<0.05	373.4	187.0

ACKNOWLEDGMENTS

Steve Kolinski generously shared his time and knowledge of the calorimeter and ashing oven, and provided valuable discussions of the methods and results. Dave Krupp provided instruction and facilities for the bomb calorimetry. Gordon Grau provided the freeze-dryer and access to his laboratory. Tom Leedom kindly assisted with use of the freeze-dryer. Evelyn Cox provided supplies for ashing and use of the oven. Bruce Woollard graciously delivered dry ice daily for all the freeze-drying, and Virginia Moriwake and Hal Richman helped maintain the dryer temperature. Paul Nachtigall and James Parrish made helpful comments on earlier versions of this manuscript. This paper is funded by a grant from the National Oceanic and Atmospheric Administration, Project No. RSM-7, which is sponsored by the University of Hawaii Sea Grant College Program, SOEST, under Institutional Grant No. NA86RG0041 from NOAA Office of Sea Grant, Department of Commerce. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies. UNIHI-SEAGRANT-JC-02-06. This is HIMB contribution 1121.

Benoit-Bird, K. J., and Au, W. W. L. (**2001**). ''Target strength measurements of animals from the Hawaiian mesopelagic boundary community,'' J. Acoust. Soc. Am. **110**, 812–819.

Bethke, E., Arrhenius, F., Cardinale, M., and Hakansson, N. (**1994**). ''Comparison of the selectivity of three pelagic sampling trawls in a hydroacoustic survey,'' Fish. Res. Amst. **41**, 15–23.

Claytor, R. R., Clay, A., Walter, E., Jorgenson, J., Clement, M., and St. Hilaire, A. (**1999**). ''National Hydroacoustic Program Client Participation projects: Fleet acoustics, partnership review, potential users, equipment inventory,'' Canadian Technical Report of Fisheries and Aquatic Sciences **2272**, 1–76.

Cummins, K. C., and Wuycheck, J. C. (**1977**). ''Caloric equivalents for investigations in ecological energetics,'' Int. Assn. Theoret. Appl. Limnology **18**, 1–158.

Donnelly, J., Stickney, D. G., and Torres, J. J. (**1993**). ''Proximate and elemental composition and energy content of mesopelagic crustaceans from the eastern Gulf of Mexico,'' Mar. Biol. (Berlin) **115**, 459–480.

Donnelly, J., Torres, J. J., Hopkins, T. L., and Lancraft, T. M. (**1990**). ''Proximate composition of Antarctic mesopelagic fish,'' Mar. Biol. (Berlin) **106**, 13–23.

Foote, K. G. (**1983**). ''Linearity of fisheries acoustics with addition theorem,'' J. Acoust. Soc. Am. **73**, 1932–1940.

Golley, F. B. (**1961**). ''Energy values of ecological materials,'' Ecology **42**, 581–584.

Ikeda, T. (**1996**). ''Metabolism, body composition, and energy budget of the mesopelagic fish *Maurolicus muelleri* in the Sea of Japan,'' Fish. Bull. **94**, 49–58.

Kemper, J. H., and Raat, A. J. P. (**1997**). ''Hydroacoustic assessment of the fish stock in Theodorushaven, a small Dutch harbour,'' Fisheries Management Ecology **4**, 63–71.

Liao, C.-H., Lee, K.-T., Lee, M.-A., and Lu, H.-J. (**1999**). ''Biomass distribution and zooplankton composition of the sound-scattering layer in the waters of southern East China Sea,'' ICES J. Mar. Sci. **56**, 766–778.

Love, R. H. (**1970**). ''Dorsal-aspect target strength of an individual fish,'' J. Acoust. Soc. Am. **49**, 816–823.

MacLennan, D. N., and Simmonds, E. J. (**1992**). *Fisheries Acoustics* (Chapman and Hall, New York).

Medwin, H., and Clay, C. (**1997**). *Fundamentals of Acoustical Oceanography* (Academic, San Diego).

Neighbors, M. A., and Nafpaktitis, B. G. (**1982**). ''Lipid compositions, water contents, swimbladder morphologies and buoyancies of 19 species of midwater fish (18 myctophids and 1 neoscopelid),'' Mar. Biol. (Berlin) **66**, 207–215.

Ohshima, T., Ratnayake, W. M. N., and Ackman, R. G. (**1987**). ''Cod lipids, solvent systems and the effect of fatty acid chain length and unsaturation on lipid class analysis by Iatroscan TLC-FID,'' J. Aqu. Ocean Can. Sci. **64**, 219–223.

Paine, R. T. (**1966**). ''Endothermy in bomb calorimetry,'' Limnol. Oceanogr. **11**, 126–129.

Paine, R. T. (**1971**). ''The measurement and application of the calorie to ecological problems,'' Annu. Rev. Ecol. Syst. **2**, 145–164.

Parkinson, E. A., Rieman, B. E., and Rudstam, L. G. (**1994**). ''Comparison of acoustic and trawl methods for estimating density and age composition of kokanee,'' Trans. Am. Fish. Soc. **123**, 841–854.

Reid, S. B., Hirota, J., Young, R. E., and Hallacher, L. E. (**1991**). ''Mesopelagic-boundary community in Hawaii: Micronekton at the interface between neritic and oceanic ecosystems,'' Mar. Biol. (Berlin) **109**, 427–440.

Sarvala, J., Salonen, K., Jarvinen, M., Aro, E., Huttula, T., Kotilainen, P., Kurki, H., Langenberg, V., Mannini, P., Peltonen, A., Plisnier, P. D., Vuorinen, I., Molsa, H., and Lindqvist, O. V. (**1999**). ''Trophic structure of Lake Tanganyika: Carbon flows in the pelagic food web,'' Hydrobiologia **407**, 149–173.

Slobodkin, L. B., and Richman, L. B. (**1961**). ''Calories/gm in species of animals,'' Nature (London) **191**, 299.

Stickney, D. G., and Torres, J. J. (**1989**). ''Proximate composition and energy content of mesopelagic fish from the eastern Gulf of Mexico,'' Mar. Biol. (Berlin) **103**, 13–24.

Nonlinear two-dimensional model for thermoacoustic engines

Mark F. Hamilton, Yurii A. Ilinskii, and Evgenia A. Zabolotskaya
Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas 78712-1063 and Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029

(Received 27 September 2001; accepted for publication 11 February 2002)

A two-dimensional model and efficient solution algorithm are developed for studying nonlinear effects in thermoacoustic engines. There is no restriction on the length or location of the stack, and the cross-sectional area of the resonator may vary with position along its axis. Reduced model equations are obtained by ordering spatial derivatives in terms of rapid variations across the pores in the stack, versus slow variations along the resonator axis. High efficiency is achieved with the solution algorithm because the stability condition for numerical integration of the model equations is connected with resonator length rather than pore diameter. Computation time is reduced accordingly, by several orders of magnitude, without sacrificing spatial resolution. The solution algorithm is described in detail, and the results are verified by comparison with established linear theory. Two examples of nonlinear effects are investigated briefly, the onset of instability through to saturation and steady state, and nonlinear waveform distortion as a function of resonator shape. [DOI: 10.1121/1.1467675]

PACS numbers: 43.35.Ud, 43.25.Gf [VWS]

I. INTRODUCTION

A variety of nonlinear acoustical effects is readily observed in thermoacoustic engines.[1–4] While various analytical models are available[5,6] (see also other investigations cited in these papers), they are based on substantial approximations and the solutions are often unwieldy. An accurate description that accounts for the combined effects of nonlinearity, viscosity, heat conduction, and multidimensional flow thus calls for numerical solution of the equations of motion. Previous numerical investigations[7–9] that account for two-dimensional motion rely on computations performed only in the neighborhood of the stack. Moreover, they are based on the assumption that the stack is short compared with the length of the resonator.

The greatest impediment to direct numerical simulation has been the computation time required for adequate resolution of the field throughout the resonator. The difficulty results from the widely different length scales encountered in thermoacoustic engines. The large dimension is the length of the resonator, which is comparable to the acoustic wavelength at the fundamental natural frequency. The small dimension is the diameter of the pores in the stack, which is comparable to the thermal penetration depth. These two lengths can differ by several orders of magnitude. For adequate resolution of the field structure within the pores, stability conditions require time steps so small compared to the period of acoustic oscillation that the computation time is prohibitive.

One solution is to develop a theoretical model that permits the stability condition to be decoupled from the small spatial scale determined by the pores. We obtain such a model by ordering spatial derivatives according to a small parameter defined by the ratio of thermal penetration depth to acoustic wavelength. A consequence of this ordering procedure is that the transverse component of the momentum equation may be ignored. The stability condition is then associated with acoustic wavelength rather than thermal penetration depth, and computation time is thus reduced by several orders of magnitude.

The model is developed for a 2D rectangular geometry that is periodic transverse to the axis of the resonator. It is assumed that the plates in the stack have negligible thickness, but there is no restriction on the length or location of the stack. The computation region is thereby reduced to a rectangular channel the length of the resonator and half the width of a single pore. Resonators with varying cross sections are described in curvilinear coordinates that transform the boundary conditions and model equations into forms nearly identical to those for constant cross section.

The model equations are obtained in Sec. II, the solution algorithm is presented in Sec. III, and the extension to resonators with varying cross sections is discussed in Sec. IV. The validity of the approach is demonstrated in Sec. V via comparison with the linear theory developed by Rott[10] and Swift.[11] Two simulations emphasizing nonlinear phenomena are presented in Sec. VI. In one, the onset of instability through to saturation and steady state is described. In the other, nonlinear waveform distortion associated with different resonator shapes is shown, and a simple physical explanation is provided.

The main purpose of the present article is to develop the model equations and describe the solution algorithm. More extensive investigations of nonlinear effects in thermoacoustic engines are postponed to later publications. An oral presentation of this work was given in April 2001 at the 1st International Workshop on Thermoacoustics in 's-Hertogenbosch, The Netherlands.[12]

II. THEORETICAL MODEL

The thermoacoustic system is described in terms of its geometry and boundary conditions, the general governing equations are presented, and a set of reduced model equa-

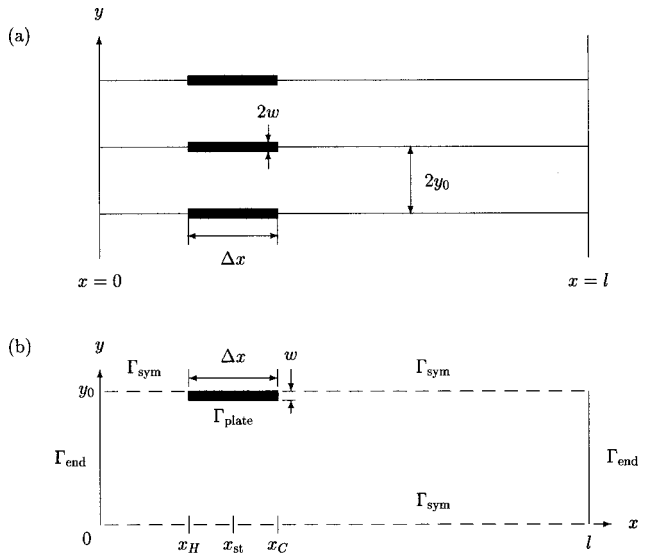


FIG. 1. (a) Geometry of engine with constant cross section and periodic properties in y direction. (b) Computation region in rectangular coordinates.

tions is obtained using a small parameter defined by the ratio of thermal penetration depth to acoustic wavelength.

A. Geometry and boundary conditions

The geometry of the thermoacoustic engine under consideration is shown in Fig. 1(a). A 2D geometry is used to eliminate variations in the z direction, and the structure is assumed to be periodic in the y direction. The spatial period is the center-to-center distance $2y_0$ between adjacent plates of thickness $2w$ in a stack of length Δx. A principal assumption is that the length scales are related as follows:

$$w \ll y_0 \ll l, \tag{1}$$

where l is the length of the resonator. These conditions permit considerable simplifications and reduction in computation time. There is no restriction on the relation of Δx to l.

We begin with the first inequality, $w \ll y_0$, and further assume that the plate thickness is negligible, such that it may be considered in the limit

$$w \to 0. \tag{2}$$

Although small but finite plate thickness may be included, it introduces complexities that are better set aside for discussion elsewhere. On the basis of Eq. (2) and periodicity, the computation region may be restricted to the narrow rectangular channel of length l and width y_0 that is shown in Fig. 1(b).

The boundary conditions imposed by symmetry are

$$\frac{\partial \phi}{\partial y}, \frac{\partial u_x}{\partial y}, u_y = 0 \quad \text{on} \quad \Gamma_{\text{sym}}, \tag{3}$$

where Γ_{sym} is the segment of the boundary designated by the dashed lines in Fig. 1(b), ϕ is any of the scalar field variables (pressure p, density ρ, temperature T), and u_i are components of the particle velocity vector. The solid surfaces of the resonator are shown as solid lines in Fig. 1(b), labeled as Γ_{end} for the ends of the resonator and Γ_{plate} for the plate. These surfaces are assumed to be rigid, and the boundary conditions on the particle velocity are thus

$$u_x = u_y = 0 \quad \text{on} \quad \Gamma_{\text{end}} \quad \text{and} \quad \Gamma_{\text{plate}}. \tag{4}$$

For the temperature we set

$$\frac{\partial T}{\partial x} = 0 \quad \text{on} \quad \Gamma_{\text{end}}, \tag{5}$$

which is a statement that no heat flows to or from the ends of the resonator. Along the plate we have

$$T = T_s \quad \text{on} \quad \Gamma_{\text{plate}}, \tag{6}$$

where T_s is the local temperature of the solid. A second condition along the plate relates $\partial T/\partial y$ to heat flow into the plate and diffusion within the plate. Since $\partial T/\partial y$ on Γ_{plate} is a source term in the heat equation for the plate, it is introduced later in this context.

A thermoacoustic refrigerator requires an external source of sound power. One means of supplying this power, and especially simple to model, is vibration of the entire resonator. This method of excitation is employed in some industrial applications involving acoustical resonators used as pumps.[13] The boundary conditions in Eqs. (3)–(6) remain the same in a coordinate system attached to the moving resonator, and one need only introduce an inertial body force in the equations of motion. If the resonator vibrates with acceleration $a(t)$ along the x axis, the required components of the body force (per unit volume) on the gas in the moving coordinate system are

$$f_x = -\rho a(t), \quad f_y = 0. \tag{7}$$

It is this method of mechanical excitation that we consider in the present paper.

Alternatively, one may wish to model excitation of the sound field by the motion of a piston at one end (or both ends) of the resonator. In this case, the resonator is at rest and thus $f_x = f_y = 0$. A moving piston in a stationary resonator constitutes a time-varying geometry that in turn requires a time-varying computation grid. To avoid this complication one can approximate the exact description of the moving piston with a suitable boundary condition in a stationary plane. The boundary condition in the stationary plane must be applied to momentum rather than velocity, i.e., $\rho u_x = g(t)$ and $\rho u_y = 0$, where $g(t)$ represents the oscillatory motion of the piston. If one applies the boundary condition to u_x alone, a dc velocity component must be included to ensure zero mass flow through the boundary. The dc flow results from defining oscillatory motion in fixed Eulerian coordinates.[14]

Hot and cold heat exchangers shall be modeled by imposing constant mean temperatures in the volume of gas outside the stack:

$$T_m = T_H, \quad 0 \leqslant x \leqslant x_H, \tag{8}$$

$$= T_C, \quad x_C \leqslant x \leqslant l, \tag{9}$$

where T_m indicates the time average of T, and T_H and T_C are the temperatures prescribed for the hot and cold sides, respectively. Implementation of these conditions is discussed in Secs. V and VI.

Variation in the cross-sectional area of the resonator (not depicted in Fig. 1) is also taken into account. This modification is postponed to Sec. IV, after the solution algorithm for a resonator with constant cross section has been described.

B. Governing equations

The fluid satisfies the following conservation equations for mass, momentum, and energy in an Eulerian coordinate system attached to the resonator:[15]

$$\frac{\partial \rho}{\partial t}+\frac{\partial(\rho u_i)}{\partial x_i}=0, \tag{10}$$

$$\frac{\partial(\rho u_i)}{\partial t}+\frac{\partial(\rho u_i u_k)}{\partial x_k}=-\frac{\partial p}{\partial x_i}+\frac{\partial \sigma_{ik}}{\partial x_k}+f_i, \tag{11}$$

$$\frac{\partial e}{\partial t}+\frac{\partial[(e+p)u_i]}{\partial x_i}=\frac{\partial}{\partial x_i}\left(\kappa\frac{\partial T}{\partial x_i}\right)+\frac{\partial(u_i\sigma_{ik})}{\partial x_k}+f_i u_i, \tag{12}$$

where

$$\sigma_{ik}=\mu\left(\frac{\partial u_i}{\partial x_k}+\frac{\partial u_k}{\partial x_i}-\frac{2}{3}\delta_{ik}\frac{\partial u_l}{\partial x_l}\right)+\mu_B\delta_{ik}\frac{\partial u_l}{\partial x_l} \tag{13}$$

is the viscous stress tensor, μ the shear viscosity, μ_B the bulk viscosity, δ_{ik} the Kronecker delta, κ the thermal conductivity, f_i the body force defined in Eqs. (7), and

$$e=\tfrac{1}{2}\rho(u_x^2+u_y^2)+\epsilon \tag{14}$$

the energy density (per unit volume). The fluid is assumed to be an ideal gas, for which the internal energy ϵ and equation of state are related as

$$\epsilon=c_v\rho T=\frac{p}{\gamma-1}, \tag{15}$$

where $\gamma=c_p/c_v$ is the ratio of the specific heats c_p and c_v at constant pressure and volume, respectively.

The motivation for expressing the left-hand sides of Eqs. (10)–(12) in conservation-law form is that finite-difference algorithms are available that guarantee conservation of the given quantities. Most important for our purposes is conservation of the total mass and energy in the resonator, especially when steady-state conditions are of interest. Since the gas in the resonator is initially at rest, computations that extend over hundreds of acoustic periods are required before all significant transient effects disappear.

The temperature distribution T_s in the plate is governed by

$$\rho_s c_s\frac{\partial T_s}{\partial t}=\frac{\partial}{\partial x}\left(\kappa_s\frac{\partial T_s}{\partial x}\right)-\frac{\kappa}{w}\frac{\partial T}{\partial y}\bigg|_{\Gamma_{\text{plate}}}, \tag{16}$$

where ρ_s and c_s are the density and specific heat of the plate, respectively, and κ_s is its thermal conductivity. No y dependence is taken into account by the diffusion term in Eq. (16) because the gradient $\partial T_s/\partial y$ vanishes in the limit of zero plate thickness. The second term on the right-hand side accounts for heat exchange with the gas in contact with both sides of the plate.

All coefficients in the preceding equations may vary as functions of position or according to values of the local state variables. For example, it is assumed in the calculations for the simulations presented in Secs. V and VI that the viscosity and thermal conductivity of the gas depend on temperature.

C. Ordering scheme

Because of the large difference in length scales encountered in thermoacoustic engines, direct numerical integration of the equations in Sec. II B is exceedingly time intensive. The small scale, in the y direction, is the pore size y_0 in the stack, which is the same order as the thermal penetration depth δ_κ in the gas. The large scale, in the x direction, is the length l of the resonator, which is the same order as the acoustic wavelength λ at the fundamental natural frequency. We thus have

$$y_0\sim\delta_\kappa, \quad l\sim\lambda. \tag{17}$$

These quantities may be used to estimate numerical step sizes required to satisfy stability criteria.

We begin by taking

$$\delta_\kappa/\lambda\sim y_0/l\sim 10^{-3}, \tag{18}$$

which is a nominal value for air at 1 kHz (this ratio is proportional to the square root of frequency). The Courant stability conditions[16] for the step sizes Δx, Δy, and Δt are $\Delta x/\Delta t>c$ and $\Delta y/\Delta t>c$, where c is the local value of the sound speed. The latter condition is the more restrictive, and taking $\Delta y/y_0\sim 10^{-2}$ (for adequate resolution of the field between the plates) we obtain

$$\Delta t<\frac{\Delta y}{c}\sim 10^{-5}\,\frac{\lambda}{c} \tag{19}$$

on the basis of Eqs. (17) and (18). With λ/c recognized to be comparable to the period of the acoustic oscillations at the fundamental natural frequency of the resonator, stability thus requires on the order of 10^5 samples per cycle of the acoustic waveform.

A method for circumventing the stability condition in the y direction, to avoid severe oversampling of the acoustic waveform, is to simplify the governing equations by introducing the small ordering parameter

$$\eta\sim\delta_\kappa/\lambda\sim y_0/l. \tag{20}$$

The relative ordering of the derivatives is thus

$$\frac{\partial}{\partial t}=O(\eta^0), \quad \frac{\partial}{\partial x}=O(\eta^0), \quad \frac{\partial}{\partial y}=O(\eta^{-1}). \tag{21}$$

From Eqs. (10) and (11) one obtains

$$\frac{u_y}{u_x}=O(\eta), \quad \frac{\partial p/\partial y}{\partial p/\partial x}=O(\eta), \tag{22}$$

which indicate that the fluid motion and pressure gradients are predominantly in the x direction. The model equations obtained on the basis of Eqs. (21) and (22) permit reduction in computation time by $O(\eta)$.

D. Model equations

Taking Eqs. (21) into account, one finds that the leading order relations following from Eqs. (10)–(16) are

$$\frac{\partial \rho}{\partial t} = -\frac{\partial(\rho u_x)}{\partial x} - \frac{\partial(\rho u_y)}{\partial y}, \tag{23}$$

$$\frac{\partial(\rho u_x)}{\partial t} = -\frac{\partial(\rho u_x^2)}{\partial x} - \frac{\partial(\rho u_x u_y)}{\partial y} - \frac{\partial p}{\partial x} + \frac{\partial}{\partial y}\left(\mu \frac{\partial u_x}{\partial y}\right) - \rho a(t), \tag{24}$$

$$\frac{\partial e}{\partial t} = -\frac{\partial[(e+p)u_x]}{\partial x} - \frac{\partial[(e+p)u_y]}{\partial y} + \frac{\partial}{\partial y}\left(\kappa \frac{\partial T}{\partial y}\right) + \frac{1}{2}\frac{\partial}{\partial y}\left(\mu \frac{\partial u_x^2}{\partial y}\right) - \rho u_x a(t), \tag{25}$$

$$\rho_s c_s \frac{\partial T_s}{\partial t} = \frac{\partial}{\partial x}\left(\kappa_s \frac{\partial T_s}{\partial x}\right) - \frac{\kappa}{w}\frac{\partial T}{\partial y}\bigg|_{\Gamma_{\text{plate}}}, \tag{26}$$

where

$$e = \tfrac{1}{2}\rho u_x^2 + \epsilon, \tag{27}$$

$$\epsilon = c_v \rho T = \frac{p}{\gamma - 1}. \tag{28}$$

Equations (23), (26), and (28) are unaltered. In Eqs. (24) and (25), only the higher-order loss terms (depending on μ and κ) are omitted. In Eq. (27), the contribution of u_y to the kinetic energy is omitted. All boundary conditions in Sec. II A remain the same.

The most noticeable omission is the entire y component of the momentum equation. Only the x component, Eq. (24), is retained. Because numerical solution of the y component is avoided, the stability condition is connected to Δx rather than Δy. The maximum time step indicated by Eq. (19) is thus reduced by several orders of magnitude.

Justification for omitting the y component of the momentum equation follows from the smallness of the pressure gradient $\partial p/\partial y$, as indicated by the second of Eqs. (22). The characteristic time required for the pressure to equilibrate across the channel in Fig. 1(b) is

$$t_{\text{eq}} \simeq y_0/c \sim \eta\lambda/c, \tag{29}$$

which is a factor of η less than one acoustic period. On the basis of this observation, we impose the condition

$$\partial p/\partial y = 0 \tag{30}$$

following each time step Δt in the solution algorithm. Equation (30) is recognized as a principal component of Prandtl's boundary layer theory for incompressible flow.[17] As shown in Sec. III C, combining Eq. (30) with (25) permits u_y to be determined from the energy equation, rather than from the y component of the momentum equation. Equations (23)–(28) and (30) thus constitute a system of eight equations in eight unknowns.

III. SOLUTION ALGORITHM

Integration of Eqs. (23)–(26) over time step Δt provides the values of ρ, ρu_x, e, and T_s required to initiate the next step. These quantities, in combination with Eqs. (27) and (28), provide the values of u_x, p, and T appearing on the right-hand sides of Eqs. (23)–(26). The remaining field variable, u_y, is determined in the third substep by making use of Eq. (30) as described below.

A splitting algorithm is used to integrate Eqs. (23)–(26) from time t_k to time $t_{k+1} = t_k + \Delta t$. Three substeps are employed, one to account for variations in x over time step Δt, and two to account for variations in y over the same time step Δt. Under the assumption of negligible plate thickness w, the spatial domain shown in Fig. 1(b) is the rectangular region defined by $0 \leqslant x \leqslant l$ and $0 \leqslant y \leqslant y_0$. All field variables except u_y are evaluated on the grid (x_i, y_j) defined by $x_i = (i - \frac{1}{2})\Delta x$ and $y_j = (j - \frac{1}{2})\Delta y$, where $i,j = 1,2,3\ldots$. The grid $(x_i, y_{j'})$ for u_y is shifted one half-step in the y direction, such that $y_{j'} = j'\Delta y$, where $j' = 0,1,2\ldots$.

A. First substep

The equations for the first substep, from k to $k + 1/3$, are the ones obtained by omitting all terms with y derivatives in Eqs. (23)–(26):

$$\frac{\partial \rho}{\partial t} = -\frac{\partial(\rho u_x)}{\partial x}, \tag{31}$$

$$\frac{\partial(\rho u_x)}{\partial t} = -\frac{\partial(\rho u_x^2 + p)}{\partial x} - \rho a(t), \tag{32}$$

$$\frac{\partial e}{\partial t} = -\frac{\partial[(e+p)u_x]}{\partial x} - \rho u_x a(t), \tag{33}$$

$$\frac{\partial T_s}{\partial t} = \frac{1}{\rho_s c_s}\frac{\partial}{\partial x}\left(\kappa_s \frac{\partial T_s}{\partial x}\right). \tag{34}$$

Equations (31)–(33) are solved with a two-step Lax–Wendroff scheme.[16] Equation (34) can be solved with either an explicit or semi-implicit scheme. The terminology semi-implicit is used because $\kappa_s(T_s)$ is evaluated at step k in the otherwise implicit finite-difference equation for determining the value of T_s at step $k + 1/3$.

For clarity and future reference, we define

$$D_n(\delta, \phi) = \left[\frac{\partial}{\partial z}\left(\delta \frac{\partial \phi}{\partial z}\right)\right]_n \tag{35}$$

to be the finite-difference operator involving the variables δ and ϕ. The formulation of this operator based on discretization $z_{n+1} = z_n + \Delta z$ is

$$D_n(\delta, \phi) = \frac{1}{2(\Delta z)^2}[(\delta_{n+1} + \delta_n)\phi_{n+1} - (\delta_{n+1} + 2\delta_n + \delta_{n-1}) \times \phi_n + (\delta_n + \delta_{n-1})\phi_{n-1}]. \tag{36}$$

For example, the operator assumes the form $D_i(\kappa_{s,k}, T_{s,k+1/3})$ in the scheme for solving Eq. (34).

B. Second substep

The second substep, from $k+1/3$ to $k+2/3$, involves solution of the equations

$$\frac{\partial(\rho u_x)}{\partial t}=\frac{\partial}{\partial y}\left(\mu\frac{\partial u_x}{\partial y}\right), \tag{37}$$

$$\frac{\partial e}{\partial t}=\frac{\partial}{\partial y}\left(\kappa\frac{\partial T}{\partial y}\right)+\frac{1}{2}\frac{\partial}{\partial y}\left(\mu\frac{\partial u_x^2}{\partial y}\right), \tag{38}$$

$$\frac{\partial T_s}{\partial t}=-\frac{\kappa}{\rho_s c_s w}\frac{\partial T}{\partial y}\bigg|_{\Gamma_{\text{plate}}}, \tag{39}$$

where the terms on the right-hand sides are all terms which were omitted in Eqs. (31)–(34) except for those which depend on u_y. Equation (39) is a boundary condition in the algorithm for solving Eq. (38).

We begin with the algorithm for integrating Eq. (37), the semi-implicit finite-difference form of which is

$$(\rho u_x)_{k+2/3}=(\rho u_x)_{k+1/3}+D_j(\mu_{k+1/3},u_{x,k+2/3})\Delta t. \tag{40}$$

Now consider the solution of the associated equation

$$\beta\frac{\partial u_x}{\partial t}=\frac{\partial}{\partial y}\left(\mu\frac{\partial u_x}{\partial y}\right), \tag{41}$$

where β is a parameter. The corresponding discretized form of Eq. (41) is

$$D_j(\mu_{k+1/3},u_{x,k+2/3})\Delta t=\beta(u_{x,k+2/3}-u_{x,k+1/3}), \tag{42}$$

substitution of which in Eq. (40) yields

$$(\rho u_x)_{k+2/3}=(\rho_{k+1/3}-\beta)u_{x,k+1/3}+\beta u_{x,k+2/3}. \tag{43}$$

Choosing $\beta=\rho_{k+1/3}$ we arrive at the desired solution of Eq. (37):

$$(\rho u_x)_{k+2/3}=\rho_{k+1/3}u_{x,k+2/3}. \tag{44}$$

The velocity $u_{x,k+2/3}$ is obtained by solving Eq. (42) with $\beta=\rho_{k+1/3}$.

A similar approach is followed to integrate Eq. (38). Consider first the variation due to heat conduction. Since heat conduction contributes only to the change in internal energy and not kinetic energy, we have from Eqs. (27) and (38)

$$\frac{\partial\epsilon}{\partial t}=\frac{\partial}{\partial y}\left(\kappa\frac{\partial T}{\partial y}\right). \tag{45}$$

The boundary conditions on the temperature are given by Eqs. (3), (5), (6), and (39). The total energy at step $k+2/3$ is

$$e_{k+2/3}=(\tfrac{1}{2}\rho u_x^2)_{k+1/3}+\epsilon_{k+2/3}, \tag{46}$$

where $\epsilon_{k+2/3}$ is obtained by solving the semi-implicit finite-difference form of Eq. (45):

$$\epsilon_{k+2/3}=\epsilon_{k+1/3}+D_j(\kappa_{k+1/3},T_{k+2/3})\Delta t. \tag{47}$$

To obtain this solution consider the associated equation

$$\beta\frac{\partial T}{\partial t}=\frac{\partial}{\partial y}\left(\kappa\frac{\partial T}{\partial y}\right), \tag{48}$$

the finite-difference form of which is

$$D_j(\kappa_{k+1/3},T_{k+2/3})\Delta t=\beta(T_{k+2/3}-T_{k+1/3}). \tag{49}$$

Now choose $\beta=c_v\rho_{k+1/3}$, substitute Eq. (49) in (47), use Eq. (28) to write $\epsilon=c_v\rho T$, and thus obtain

$$\epsilon_{k+2/3}=c_v\rho_{k+1/3}T_{k+2/3}. \tag{50}$$

We must still include the contribution due to viscosity in Eq. (38). It is sufficient to perform this calculation with an explicit scheme, the result of which simply adds an energy increment equal to $\frac{1}{2}D_j(\mu_{k+1/3},u_{x,k+1/3}^2)\Delta t$. Including this increment together with Eq. (50) in (46) yields

$$e_{k+2/3}=(\tfrac{1}{2}\rho u_x^2)_{k+1/3}+c_v\rho_{k+1/3}T_{k+2/3}+\tfrac{1}{2}D_j(\mu_{\kappa+1/3},u_{x,k+1/3}^2)\Delta t. \tag{51}$$

This is the desired solution of Eq. (38), where $T_{k+2/3}$ is obtained by solving Eq. (49) with $\beta=c_v\rho_{k+1/3}$.

C. Third substep

The third and final substep, from $k+2/3$ to $k+1$, requires solution of the set

$$\frac{\partial\rho}{\partial t}=-\frac{\partial(\rho u_y)}{\partial y}, \tag{52}$$

$$\frac{\partial(\rho u_x)}{\partial t}=-\frac{\partial(\rho u_x u_y)}{\partial y}, \tag{53}$$

$$\frac{\partial(\rho u_x^2)}{\partial t}=-\frac{\partial(\rho u_x^2 u_y)}{\partial y}, \tag{54}$$

$$\frac{\partial\epsilon}{\partial t}=-\frac{\partial[(\epsilon+p)u_y]}{\partial y}, \tag{55}$$

which accounts for all remaining terms in Eqs. (23)–(26). Note that the energy terms have been separated into Eq. (54) for the kinetic energy and Eq. (55) for the internal energy.

To proceed we need the value of u_y, which is determined by applying the condition $\partial p/\partial y=0$ from Eq. (30) to Eq. (55). Using the relation $\epsilon=p/(\gamma-1)$ we thus rewrite Eq. (55) as

$$\frac{\partial p}{\partial t}=-\gamma\frac{\partial(pu_y)}{\partial y}=-\gamma p\frac{\partial u_y}{\partial y}, \tag{56}$$

where the second equality follows from $\partial p/\partial y=0$, and the implicit finite-difference form is

$$\frac{p_{k+1}^j-p_{k+2/3}^j}{\Delta t}=-\gamma p_{k+1}^j\frac{u_{y,k+1}^{j'}-u_{y,k+1}^{j'-1}}{\Delta y}, \tag{57}$$

where $j'=j$. We again invoke $\partial p/\partial y=0$ by imposing the condition that at the end of the third substep, the pressure equilibrates across the channel to the average of its value at the beginning of the substep:

$$p_{k+1}\equiv\langle p_{k+2/3}^j\rangle_j=\frac{1}{N_y}\sum_{j=1}^{N_y}p_{k+2/3}^j, \tag{58}$$

which is independent of y and therefore of index j. Here, N_y is the number of points in the y direction, and $\langle\cdot\rangle_j$ indicates the average of a quantity across the channel. Equation (57) may now be rearranged as

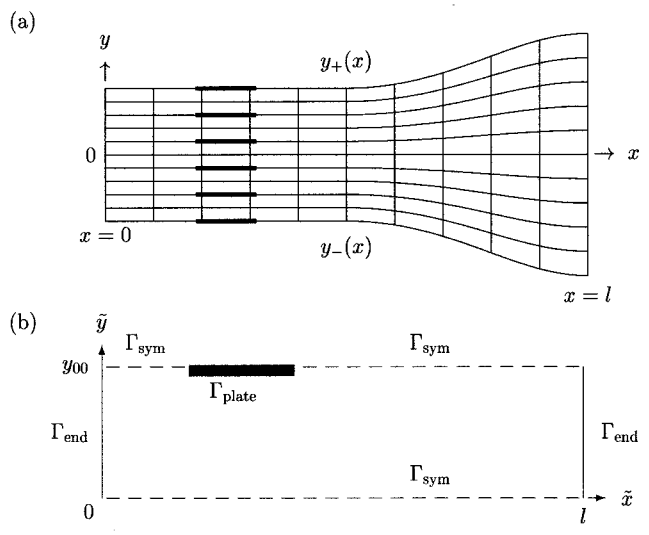


FIG. 2. (a) Geometry of engine with varying cross section. The mesh displays constant values of the curvilinear coordinates. (b) Computation region in curvilinear coordinates.

$$u_{y,k+1}^{j'}=u_{y,k+1}^{j'-1}-\left(1-\frac{p_{k+2/3}^{j}}{\langle p_{k+2/3}^{j}\rangle_j}\right)\frac{\Delta y}{\gamma\Delta t}, \tag{59}$$

which is the algorithm for calculating u_y across the channel. Physically, it determines the transverse velocity required to eliminate the transverse pressure gradients developed in the first two substeps.

The final pressure $p_{k+1}^{i,j}$ (and therefore internal energy $\epsilon_{k+1}^{i,j}$) and transverse velocity $u_{y,k+1}^{i,j'}$ are thus determined by Eqs. (58) and (59), respectively. The final values of the density $\rho_{k+1}^{i,j}$, momentum $(\rho u_x)_{k+1}^{i,j}$, and kinetic energy $(\frac{1}{2}\rho u_x^2)_{k+1}^{i,j}$ are obtained from explicit finite-difference solutions of Eqs. (52), (53), and (54), respectively.

The third substep ensures that the condition $\partial p/\partial y=0$ is maintained by the conservation equations. Transverse pressure gradients are generated mainly in the second substep, and we therefore set $\partial p/\partial y=0$ after completion of the second substep. A significant transverse pressure gradient is also established in the first substep, although only at the ends of the stack, $x=x_H$ and $x=x_C$. The procedure described for the third substep is thus applied twice, not only after the second substep, but also between the two half-steps in the Lax–Wendroff scheme.

IV. VARIABLE CROSS SECTION

We describe here the modifications in the algorithm required to account for a resonator whose cross-sectional area varies as $S(x)=w_y(x)w_z$, where $w_y(x)$ is the varying width in the y direction, and w_z is a constant width in the z direction. As shown in Fig. 2(a), the geometry is assumed to be symmetric about the plane $y=0$. The upper and lower boundaries of the resonator are defined by $y_\pm=\pm\frac{1}{2}w_y(x)$, or

$$y_\pm=\pm\tfrac{1}{2}w_0 s(x), \tag{60}$$

where $s(x)=S(x)/S_0$ is a dimensionless area, S_0 a reference area, and $w_0=S_0/w_z$ a reference width in the y direction. The separation distance between adjacent plates in the stack is assumed to vary uniformly with x and to be given by $2y_0(x)=2y_{00}s(x)$, where y_{00} is a reference separation distance.

We employ a curvilinear coordinate system $(\tilde{x},\tilde{y})$ defined by

$$\tilde{x}=x, \quad \tilde{y}=y/s(x), \tag{61}$$

such that the nonrectangular geometry in the Cartesian system (x,y) is mapped into a rectangular geometry. Constant values of $\tilde{x}$ and $\tilde{y}$ are displayed in Fig. 2(a) as vertical and horizontal lines, respectively. As shown in the Appendix, the corresponding transformation of Eqs. (23)–(26) yields

$$\frac{\partial\tilde{\rho}}{\partial t}=-\frac{\partial(\tilde{\rho}\tilde{u}_x)}{\partial\tilde{x}}-\frac{\partial(\tilde{\rho}\tilde{u}_y)}{\partial\tilde{y}}, \tag{62}$$

$$\frac{\partial(\tilde{\rho}\tilde{u}_x)}{\partial t}=-\frac{\partial(\tilde{\rho}\tilde{u}_x^2)}{\partial\tilde{x}}-\frac{\partial(\tilde{\rho}\tilde{u}_x\tilde{u}_y)}{\partial\tilde{y}}-\frac{\partial\tilde{p}}{\partial\tilde{x}}+\frac{s'}{s}\tilde{p}+\frac{1}{s}\frac{\partial}{\partial\tilde{y}}\left(\mu\frac{\partial\tilde{u}_x}{\partial\tilde{y}}\right)-\tilde{\rho}a(t), \tag{63}$$

$$\frac{\partial\tilde{e}}{\partial t}=-\frac{\partial[(\tilde{e}+\tilde{p})\tilde{u}_x]}{\partial\tilde{x}}-\frac{\partial[(\tilde{e}+\tilde{p})\tilde{u}_y]}{\partial\tilde{y}}+\frac{1}{s}\frac{\partial}{\partial\tilde{y}}\left(\kappa\frac{\partial\tilde{T}}{\partial\tilde{y}}\right)+\frac{1}{2s}\frac{\partial}{\partial\tilde{y}}\left(\mu\frac{\partial\tilde{u}_x^2}{\partial\tilde{y}}\right)-\tilde{\rho}\tilde{u}_x a(t), \tag{64}$$

$$\frac{\partial\tilde{T}_s}{\partial t}=\frac{1}{\rho_s c_s}\frac{\partial}{\partial\tilde{x}}\left(\kappa_s\frac{\partial\tilde{T}_s}{\partial\tilde{x}}\right)-\frac{\kappa}{\rho_s c_s w s}\left.\frac{\partial\tilde{T}}{\partial\tilde{y}}\right|_{\Gamma_{plate}}, \tag{65}$$

and $\partial\tilde{p}/\partial\tilde{y}=0$, where

$$\tilde{\rho}=s\rho, \quad \tilde{p}=sp, \quad \tilde{e}=se, \quad \tilde{T}=T, \quad \tilde{T}_s=T_s, \tag{66}$$

and the particle velocity components are related according to

$$\tilde{u}_x=u_x, \quad \tilde{u}_y=-(s'y/s^2)u_x+u_y/s, \tag{67}$$

$$u_x=\tilde{u}_x, \quad u_y=s'\tilde{y}\tilde{u}_x+s\tilde{u}_y, \tag{68}$$

where primes indicate derivatives. Equations (62) and (64) are obtained from Eqs. (23) and (25) without approximation. A term involving $\partial\tilde{p}/\partial\tilde{y}$ was ignored in Eq. (63) on the basis of Eq. (30), and Eq. (65) was simplified by ignoring terms of relative order η^2. If the y dependence is ignored and if μ and $a(t)$ are set to zero, the linear forms of Eqs. (62) and (63), combined with an equation of state, yield the Webster horn equation[18] after the untransformed variables are reinstated.

To transform the boundary conditions it is necessary to consider how a varying cross section alters the geometry of the individual channels of width $y_0(x)$. From Fig. 2(a) it is clear that the channels are not identical. Instead, they vary with distance from the central axis of the resonator, and the structure is not periodic. However, the premise of our approach is that the area variation $s(x)$ is gradual, such that channels in the neighborhood of the central axis are similar and therefore reasonably periodic. In this case, solutions obtained for the central channel resemble closely the solutions for neighboring channels. Computations are thus performed

for the central channel alone, and the solutions are presumed to be accurate throughout a central region of the resonator containing many channels.

The computation region in transformed coordinates is the central rectangular channel defined by $0 \leqslant \tilde{x} \leqslant l$ and $0 \leqslant \tilde{y} \leqslant y_{00}$, as shown in Fig. 2(b). As before, the ends of the resonator are taken to be rigid in the planes $\tilde{x}=0$, l. To an accuracy consistent with the use of Eqs. (62)–(65), the boundary conditions given by Eqs. (3)–(6) retain the same form after transformation, specifically,

$$\frac{\partial \tilde{\phi}}{\partial \tilde{y}}, \frac{\partial \tilde{u}_x}{\partial \tilde{y}}, \tilde{u}_y = 0 \quad \text{on} \quad \Gamma_{\text{sym}}, \tag{69}$$

$$\tilde{u}_x = \tilde{u}_y = 0 \quad \text{on} \quad \Gamma_{\text{end}} \text{ and } \Gamma_{\text{plate}}, \tag{70}$$

$$\frac{\partial \tilde{T}}{\partial \tilde{x}} = 0 \quad \text{on} \quad \Gamma_{\text{end}}, \tag{71}$$

$$\tilde{T} = \tilde{T}_s \quad \text{on} \quad \Gamma_{\text{plate}}. \tag{72}$$

Equations (70) and (72) are exact. Although Eqs. (69) are exact along the central axis $\tilde{y}=0$, they are not at $\tilde{y}=y_{00}$. Errors in the transformed conditions at the upper boundary are associated with the absence of periodicity for a nonrectangular resonator. An accurate assessment of the errors requires an alternative form of solution, but an estimate may be made as follows. One may anticipate that Eqs. (69) differ from the true field condition at the upper boundary by a factor comparable to the cosine of the angle formed by the curve $y=y_0(x)$ and the x axis. This factor is approximately $[1-(dy/dx)^2]^{1/2}$, where $dy/dx = y_{00}s' \sim y_{00}/l \sim \eta$. The errors are thus $O(\eta^2)$ and may be ignored. Finally, Eq. (71) follows from Eq. (5) by ignoring the $\tilde{y}$ term in the first of Eqs. (A2).

Equations (62)–(65) are solved using the algorithm described in Sec. III, modified slightly to include the factors of s. The only new term, $s'\tilde{p}/s$ in Eq. (63), is included in the first substep.

V. COMPARISON WITH LINEAR THEORY

No suitable nonlinear theory is available as a benchmark for comparison with results obtained from the present model. However, the novel features of the solution algorithm concern approximations introduced to efficiently yet accurately resolve the 2D structure of a field having considerably different length scales. The fact that the equations to be solved are nonlinear presents no significant additional complications, and therefore comparison with linear theory should be sufficient to assess the validity of the algorithm. The linear theory for thermoacoustic engines developed by Rott[10] and Swift[11] provides an appropriate benchmark for this purpose.

In the theory presented by Rott,[10] 2D equations of motion are reduced via analytical integrations to a 1D system that is integrated numerically. His result, including the extension by Swift,[11] may be expressed as the following pair of coupled equations for the complex acoustic pressure p_1 and axial particle velocity $\langle u_1 \rangle$ associated with time dependence $e^{i\omega t}$:

$$\frac{dp_1}{dx} = -\frac{i\omega\rho_m \langle u_1 \rangle}{1-f_\mu} - \rho_m A, \tag{73}$$

$$\frac{d\langle u_1 \rangle}{dx} = -\left(1 + \frac{(\gamma-1)f_\kappa}{1+\epsilon_s}\right)\frac{i\omega p_1}{\rho_m c_m^2} + \frac{1}{T_m}\frac{dT_m}{dx}\frac{(f_\kappa - f_\mu)\langle u_1 \rangle}{(1-\text{Pr})(1+\epsilon_s)(1-f_\kappa)}, \tag{74}$$

where

$$f_{\kappa,\mu} = \frac{\tanh[(1+i)y_0/\delta_{\kappa,\mu}]}{(1+i)y_0/\delta_{\kappa,\mu}}, \quad \epsilon_s = \frac{\rho_m c_p y_0}{\rho_s c_s w} f_\kappa, \tag{75}$$

$\text{Pr} = c_p\mu/\kappa$ is the Prandtl number, A is the complex acceleration amplitude of the resonator, and the subscript m indicates the local mean (time-averaged) value of a quantity. The thermal and viscous penetration depths are defined by $\delta_\kappa = \delta_\mu/\sqrt{\text{Pr}}$ and $\delta_\mu = \sqrt{2\mu/\omega\rho_m}$, respectively, and $c_m = \sqrt{\gamma R T_m}$ is the small-signal adiabatic sound speed, where $R = c_p - c_v$. We later employ the subscript 0 to indicate the ambient value of a quantity when the gas is at rest and before a temperature difference is imposed across the stack.

With $A=0$, elimination of $\langle u_1 \rangle$ in favor of p_1 in Eqs. (73) and (74) yields Eq. (54) of Swift,[11] but with ϵ_s evaluated in the limit of negligible plate thickness compared with the thermal penetration depth in the solid. Equations (73) and (74) are obtained by averaging across the channel, and $\langle u_1 \rangle$ is thus a normalized volume velocity; see Eqs. (72) and (73) of Swift.[11] The dependence of u_1 and all other field variables on y is determined by substituting the solution for p_1 in explicit analytical expressions.[11] We note that the condition $\partial p_1/\partial y = 0$, an assumption in our theoretical model, is also an element of Rott's theory.

Equations (73) and (74) were integrated numerically with a Runge–Kutta algorithm both inside and outside (where $f_{\kappa,\mu}=0$) the stack. The boundary conditions at the ends of the stack are continuity of p_1 and $\langle u_1 \rangle$, with $\langle u_1 \rangle = 0$ at the ends of the resonator. An engine of constant cross section is assumed, with $x_{\text{st}}/l = 0.25$, $\Delta x_{\text{st}}/l = 0.125$, and $y_0/l = 0.5\times10^{-3}$. Properties of the gas are defined by $\gamma = 1.67$, $\text{Pr} = 0.67$, and $\kappa/\rho_0 c_0 c_p l = 10^{-7}$, corresponding to $\delta_\kappa/y_0 \simeq 0.50$ at resonance. The heat capacity of the stack is chosen to be sufficiently large that $\epsilon_s \ll 1$.

Power is supplied both thermally and mechanically. A mean temperature $T_m(x)$ is imposed across the stack according to Eqs. (8) and (9), with $T_C = T_0$, $T_H/T_C = 1.5$, and a constant gradient between x_H and x_C. This temperature gradient is insufficient to sustain acoustic oscillations. Instead, the oscillations are sustained by vibrating the resonator with axial displacement $\xi = \Xi e^{i\omega t}$, where $\Xi/l = 0.5\times10^{-4}$, and $\omega l/\pi c_0 = 1.0035$ in steady state. The drive frequency ω matches the lowest natural frequency of the resonator (determined numerically) in the presence of the stack and temperature gradient. Solutions of Eqs. (73) and (74) evaluated in the center of the stack, $x = x_{\text{st}}$, yield the rms values of the axial particle velocity (solid lines) and temperature (dashed lines) shown in Fig. 3.

The solution algorithm described in Sec. III was executed for the same conditions. Whereas Eqs. (73) and (74)

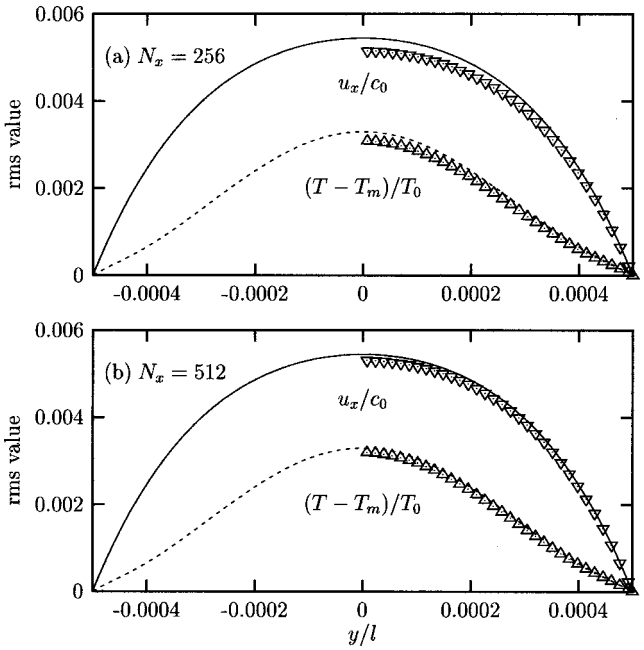


FIG. 3. Comparison of computations obtained using the present algorithm (triangles) with those obtained using the linear theory of Rott (Ref. 10) and Swift (Ref. 11) (lines) for the axial particle velocity (solid) and temperature (dashed) across a pore at the center of the stack. Two discretizations were used in the present algorithm: (a) $N_x=256$; (b) $N_x=512$.

apply to steady state, the algorithm begins with the system at rest. When the external drive is applied, the amplitude of the acoustic oscillations increases until steady state is achieved after a number of cycles corresponding to the quality factor of the system. For the given conditions, steady state was reached after about 300 cycles. Also, whereas the mean temperature distribution $T_m(x)$ could be regarded as an external parameter in the solution of Eqs. (73) and (74), it is incorporated as an initial condition in the solution algorithm. The large heat capacity prescribed for the stack ensured that this distribution was maintained until steady state was achieved.

Results from the solution algorithm are presented as triangles in Fig. 3, with $N_x=256$ evenly spaced steps for $0 \leqslant x \leqslant l$ in Fig. 3(a), $N_x=512$ in Fig. 3(b). Discretization in the transverse direction is the same for both cases, $N_y=32$ steps for $0 \leqslant y \leqslant y_0$. The number of time samples per period, N_t, is increased in proportion to N_x in accordance with the stability condition, with $N_t=1024$ and $N_t=2048$ in the two cases under consideration. Further increase in N_x provides increasingly better agreement with linear theory, while doubling N_y provides negligible improvement. We consider the quantitative agreement in Fig. 3(b) to be sufficient for our purposes here and use $N_x=512$, $N_y=32$, and $N_t=2048$ for the calculations in Sec. VI.

VI. NONLINEAR EFFECTS

We consider here the nonlinear response of a thermoacoustic prime mover, beginning with the onset of the instability[19] through to steady state. The geometry of the engine and the properties of the gas are the same as those in Sec. V except that the stack possesses a finite heat capacity determined by $\rho_s c_s w/\rho_m c_p y_0=50$, and the resonator is stationary, i.e., $a(t)=0$.

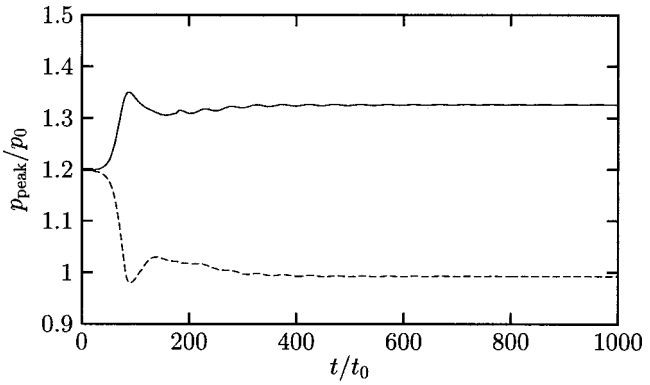


FIG. 4. Onset of oscillations and transition to steady state in an engine with constant cross section. Solid line is peak positive pressure in the waveform at the left end of the engine, dashed line is the corresponding peak negative pressure.

To maintain the temperature distribution across the stack described by Eqs. (8) and (9) requires heat exchange. For simplicity, in the present example we model virtual heat exchangers that occupy the entire volume of fluid outside the stack and present no resistance to the flow. Specifically, heat Q is simply added to or subtracted from the gas at a rate proportional to the deviation of the mean temperature T_m from the desired temperature T_H or T_C to the left or right of the stack, respectively:

$$\dot{Q}=-K(T_m-T_{H,C}). \tag{76}$$

The constant K is chosen sufficiently large that T_m is maintained reasonably close to the desired temperatures, but not so large that the sound field experiences significant attenuation. Equation (76) is included on the right-hand side of Eq. (25) [or alternatively, $s\dot{Q}$ is included on the right-hand side of Eq. (64)].

Before a temperature drop is imposed across the stack, the gas is at rest and at uniform temperature T_0, pressure p_0, and density ρ_0. At $t=0$ the temperature distribution described by Eqs. (8) and (9) is introduced, with a linear transition from x_H to x_C. For the engine under consideration, the critical temperature ratio above which acoustic oscillations are sustained is $T_H/T_C \simeq 2$. Here, we set $T_H/T_C=2.6$, where $T_C=T_0$. At this instant, the ambient pressure in the resonator increases by 20%, as indicated in Fig. 4 at $t=0$. Finally, to initiate the acoustic oscillations, a small time-varying disturbance is introduced to trigger the instability.

Figure 4 displays the peak positive (solid line) and negative (dashed line) acoustic pressure in the waveform at $x=0$ as a function of time. An exponential increase in amplitude associated with the thermoacoustic instability is observed for $t/t_0<100$, where t_0 is the fundamental period of the acoustic oscillations. Energy transfer to higher harmonics provides sufficient losses for saturation to occur at $t/t_0 \simeq 100$, and steady state is ultimately achieved for $t/t_0 \gtrsim 300$. Karpov and Prosperetti[5] describe a perturbation approach, based on a nonlinear form of Rott's 1D model, for investigating saturation of the thermoacoustic instability.

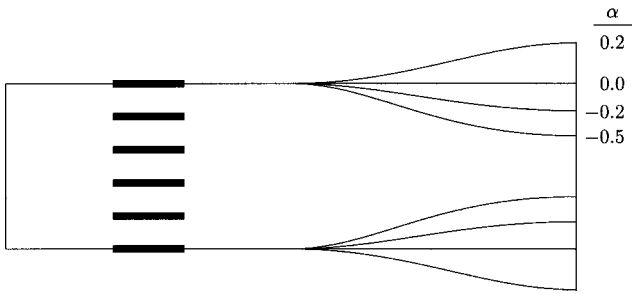


FIG. 5. Resonator shapes associated with Fig. 6.

Comparison of the positive and negative acoustic pressures in Fig. 4 shows that the waveform is asymmetric, with the negative phase having larger amplitude than the positive phase. Asymmetry occurs when the natural frequencies of the resonator are not integer multiples of one another. Because of viscous flow through the stack, the natural frequencies depart from the values $\omega_{0n}=n\pi c_0/l$ for an empty cylindrical resonator with rigid ends.

One may selectively shift the natural frequencies by varying the cross section of the resonator. We thus consider a resonator shape defined by

$$s(x)=1, \quad 0\leqslant x/l\leqslant 0.5,$$

$$=\exp\{\alpha[1+\cos(2\pi x/l)]\}, \quad 0.5\leqslant x/l\leqslant 1, \qquad (77)$$

which is depicted in Fig. 5 for several values of α. This shape function was chosen because for an empty resonator it yields a simple relation for the shift $\delta\omega_n$ of the nth natural frequency, specifically,[20] $\delta\omega_1/\omega_{01}\simeq-\alpha/4$, and $\delta\omega_n\simeq 0$ for $n\neq 1$. When the resonator is driven at its lowest natural frequency with $\alpha\neq 0$, harmonics of the drive frequency typically do not coincide with other natural frequencies of the resonator.

Shown in Fig. 6 are the waveforms at $x=0$ (solid lines) and $x=l$ (dashed lines) evaluated at $t/t_0=1000$ for the resonator shapes in Fig. 5, with all other parameters the same as in the calculations for Fig. 4. The case $\alpha=0$ corresponds exactly to Fig. 4. The general trend is that the waveforms are nominally U-shaped for $\alpha<-0.2$ and V-shaped for $\alpha>-0.2$. For $\alpha\simeq-0.2$, the resonator shape variation counteracts the natural frequency shifts introduced by the stack in such a way that $\omega_2\simeq 2\omega_1$, which provides phase-matching conditions for the fundamental and second harmonic components. Nonlinear distortion is most pronounced in this situation, and the amplitude is reduced on account of losses at the shocks.

The transition from U-shaped to V-shaped waveforms can be explained with a simple model. Let the sound pressures in the fundamental and second modes be $p_n=P_n(t)\phi_n(x)$, where ϕ_n are the normal-mode functions predicted by linear theory. We assume small perturbations of the mode shapes due to shape variations and other effects, such that $\phi_n\simeq\cos(n\pi x/l)$ for a resonator with rigid ends. If the resonator is excited in its fundamental mode, then from Eq. (90) in Ref. 20 one obtains the following relation for the lowest-order nonlinear interaction producing sound in the second mode:

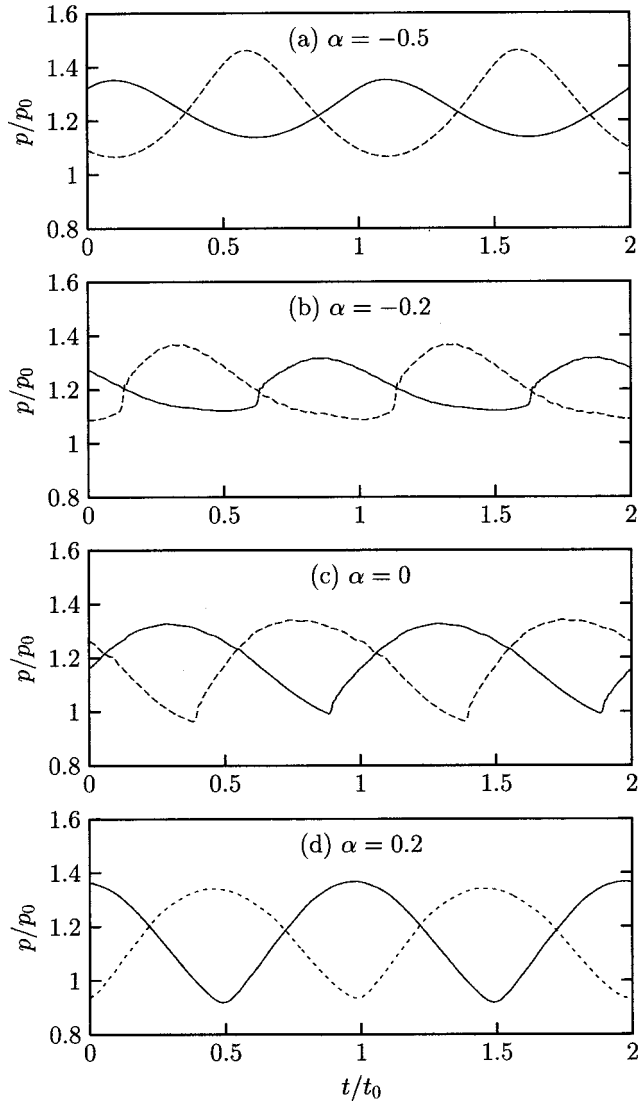


FIG. 6. Pressure waveforms at the left (solid lines) and right (dashed lines) ends of the engines depicted in Fig. 5.

$$\ddot{P}_2+2\delta_2\dot{P}_2+\omega_2^2P_2=-\frac{(\gamma+1)\pi^2}{\rho_0 l^2}P_1^2, \qquad (78)$$

where ω_n is the natural frequency and δ_n the damping coefficient of mode n. Thermoacoustic generation of sound occurs mainly in the fundamental mode at the corresponding natural frequency, so we take $P_1=P_0\cos\omega_1 t$. At $x=0$ we have $\phi_n\simeq 1$, $p(t)\simeq P_1+P_2$, and thus from Eq. (78)

$$p\simeq P_0\cos\omega_1 t+\frac{\Lambda}{4\omega_1^2}\cos 2\omega_1 t, \quad 2\omega_1>\omega_2, \qquad (79)$$

$$\simeq P_0\cos\omega_1 t-\frac{\Lambda}{2\omega_1\delta_2}\sin 2\omega_1 t, \quad 2\omega_1\simeq\omega_2, \qquad (80)$$

$$\simeq P_0\cos\omega_1 t-\frac{\Lambda}{\omega_2^2}\cos 2\omega_1 t, \quad 2\omega_1<\omega_2, \qquad (81)$$

where $\Lambda=(\gamma+1)\pi^2P_0^2/2\rho_0 l^2$. The cases $\alpha<-0.2$, $\alpha\simeq-0.2$, and $\alpha>-0.2$ correspond to Eqs. (79), (80), and (81), respectively. Indeed, the waveform described by Eq. (79) is U-shaped, and by Eq. (81) is V-shaped. The difference

is due to the π radians phase shift of the second harmonic, corresponding to the response of a simple harmonic oscillator driven above or below its natural frequency. Equation (80) reveals waveform steepening consistent with the case $\alpha \simeq -0.2$, for which positive acoustic pressures advance on negative pressures.

VII. CONCLUSION

The proposed theoretical model should prove useful for investigating fundamental physical processes and especially nonlinear effects in thermoacoustic engines. The solution algorithm makes it possible to simulate complex phenomena throughout the entire resonator and stack within reasonable computation times, yet without sacrificing spatial resolution. For example, detailed circulatory structures of acoustic streaming patterns generated inside the stack and extending into the volume of the resonator have been described with this model.[21] These and other results provide subjects for future investigations.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research, and by the Internal Research and Development Program at Applied Research Laboratories, The University of Texas at Austin.

APPENDIX: CURVILINEAR COORDINATES

Here, we rewrite the reduced Eqs. (23)–(26) in the curvilinear coordinate system given by Eqs. (61):

$$\tilde{x}=x, \quad \tilde{y}=y/s(x). \tag{A1}$$

Derivatives are transformed according to

$$\frac{\partial}{\partial x}=\frac{\partial}{\partial \tilde{x}}-\frac{s'\tilde{y}}{s}\frac{\partial}{\partial \tilde{y}}, \quad \frac{\partial}{\partial y}=\frac{1}{s}\frac{\partial}{\partial \tilde{y}}, \tag{A2}$$

where $s'=ds/dx=ds/d\tilde{x}$. Note that each of Eqs. (23)–(25) contains a divergence in the form $\partial(\phi V_i)/\partial x_i$, where ϕ is a scalar function and V_i a vector. A useful property for arbitrary coordinate systems, not necessarily orthogonal, is that the following quantity is invariant:[22]

$$\frac{1}{J}\frac{\partial(J\phi \tilde{V}_i)}{\partial \tilde{x}_i}, \tag{A3}$$

where J is the Jacobian, which for the transformation in Eq. (A1) is

$$J=\frac{\partial(x,y)}{\partial(\tilde{x},\tilde{y})}=\begin{vmatrix} 1 & 0 \\ s'\tilde{y} & s \end{vmatrix}=s. \tag{A4}$$

Now let V_i designate the particle velocity, the transformation of which between the rectangular and curvilinear coordinates is

$$\tilde{u}_i=(\partial \tilde{x}_i/\partial x_j)u_j, \quad u_i=(\partial x_i/\partial \tilde{x}_j)\tilde{u}_j, \tag{A5}$$

and from Eq. (A1) we have

$$\tilde{u}_x=u_x, \quad \tilde{u}_y=-\frac{s'y}{s^2}u_x+\frac{1}{s}u_y, \tag{A6}$$

$$u_x=\tilde{u}_x, \quad u_y=s'\tilde{y}\tilde{u}_x+s\tilde{u}_y. \tag{A7}$$

The x component of the particle velocity is thus the same in both coordinate systems. From Eqs. (A3) and (A4) we obtain the desired relation

$$\frac{\partial(\phi u_i)}{\partial x_i}=\frac{1}{s}\frac{\partial(s\phi \tilde{u}_i)}{\partial \tilde{x}_i}, \tag{A8}$$

which permits direct transformation of the divergence appearing in the conservation equations.

We begin with the continuity equation, Eq. (23), and use Eq. (A8) to write

$$\frac{\partial \rho}{\partial t}=-\frac{\partial(\rho u_i)}{\partial x_i}=-\frac{1}{s}\frac{\partial(s\rho \tilde{u}_i)}{\partial \tilde{x}_i}. \tag{A9}$$

Introduction of the transformed density $\tilde{\rho}=J\rho=s\rho$ permits Eq. (A9) to be written as

$$\frac{\partial \tilde{\rho}}{\partial t}=-\frac{\partial(\tilde{\rho}\tilde{u}_i)}{\partial \tilde{x}_i}, \tag{A10}$$

which is identical in form to Eq. (23) in rectangular coordinates.

The momentum equation, Eq. (24), is

$$\frac{\partial(\rho u_x)}{\partial t}=-\frac{\partial(\rho u_x u_i)}{\partial x_i}-\frac{\partial p}{\partial x}+\frac{\partial}{\partial y}\left(\mu\frac{\partial u_x}{\partial y}\right)-\rho a(t). \tag{A11}$$

From Eq. (A8) we have

$$\frac{\partial(\rho u_x u_i)}{\partial x_i}=\frac{1}{s}\frac{\partial(s\rho u_x \tilde{u}_i)}{\partial x_i}=\frac{1}{s}\frac{\partial(\tilde{\rho}\tilde{u}_x\tilde{u}_i)}{\partial x_i}, \tag{A12}$$

where the second equality follows from $\tilde{\rho}=s\rho$ and $\tilde{u}_x=u_x$. Equation (A11) thus becomes

$$\frac{\partial(\tilde{\rho}\tilde{u}_x)}{\partial t}=-\frac{\partial(\tilde{\rho}\tilde{u}_x\tilde{u}_i)}{\partial \tilde{x}_i}-s\frac{\partial p}{\partial x}+\frac{1}{s}\frac{\partial}{\partial \tilde{y}}\left(\mu\frac{\partial \tilde{u}_x}{\partial \tilde{y}}\right)-\tilde{\rho}a(t), \tag{A13}$$

where the viscosity term was transformed via Eqs. (A2). For the pressure term we have, from Eqs. (A2),

$$s\frac{\partial p}{\partial x}=s\frac{\partial p}{\partial \tilde{x}}-s'\tilde{y}\frac{\partial p}{\partial \tilde{y}}\simeq s\frac{\partial p}{\partial \tilde{x}}, \tag{A14}$$

where the approximation results from taking Eq. (30) into account and setting $\partial \tilde{p}/\partial \tilde{y}=0$. Letting $\tilde{p}=sp$, substitution of Eq. (A14) into (A13) yields

$$\frac{\partial(\tilde{\rho}\tilde{u}_x)}{\partial t}=-\frac{\partial(\tilde{\rho}\tilde{u}_x\tilde{u}_i)}{\partial \tilde{x}_i}-\frac{\partial \tilde{p}}{\partial \tilde{x}}+\frac{s'}{s}\tilde{p}+\frac{1}{s}\frac{\partial}{\partial \tilde{y}}\left(\mu\frac{\partial \tilde{u}_x}{\partial \tilde{y}}\right)-\tilde{\rho}a(t). \tag{A15}$$

No transformation of the resonator acceleration $a(t)$ is required because it is a kinematic quantity with a component in only the $x=\tilde{x}$ direction.

The procedure for transforming the energy equation, Eq. (25), is no different from that just described for the momentum equation, and therefore we merely present the result:

$$\frac{\partial \tilde{e}}{\partial t} = -\frac{\partial[(\tilde{e}+\tilde{p})\tilde{u}_i]}{\partial \tilde{x}_i} + \frac{1}{s}\frac{\partial}{\partial \tilde{y}}\left(\kappa \frac{\partial \tilde{T}}{\partial \tilde{y}}\right) + \frac{1}{2s}\frac{\partial}{\partial \tilde{y}}\left(\mu \frac{\partial \tilde{u}_x^2}{\partial \tilde{y}}\right) - \tilde{\rho}\tilde{u}_x a(t), \tag{A16}$$

where $\tilde{e}=se$ and $\tilde{T}=T$.

In place of Eq. (26), the heat equation for the plate is

$$\rho_s c_s \frac{\partial T_s}{\partial t} = \frac{\partial}{\partial \zeta}\left(\kappa_s \frac{\partial T_s}{\partial \zeta}\right) - \frac{\kappa}{w}\frac{\partial T}{\partial n}\bigg|_{\Gamma_{\text{plate}}}, \tag{A17}$$

where ζ is a coordinate along the plate (i.e., following a curve $\tilde{y}=\text{const}$), and n is a coordinate normal to the plate. Transformation of the derivatives yields

$$\frac{\partial}{\partial \zeta} = \frac{1}{\sqrt{1+(s'\tilde{y})^2}}\frac{\partial}{\partial \tilde{x}}, \tag{A18}$$

$$\frac{\partial}{\partial n} = \sqrt{1+(s'\tilde{y})^2}\frac{1}{s}\frac{\partial}{\partial \tilde{y}} - \frac{s'\tilde{y}}{\sqrt{1+(s'\tilde{y})^2}}\frac{\partial}{\partial \tilde{x}}. \tag{A19}$$

For plates near the center of the resonator we have $\tilde{y}\sim y_0$, such that $s'\tilde{y}=O(\eta)$ [recall Eq. (20)]. Also, $\partial/\partial\tilde{x}$ is $O(\eta)$ relative to $\partial/\partial\tilde{y}$ [recall Eq. (21)]. With $\tilde{T}_s=T_s$, Eq. (A17) thus becomes

$$\rho_s c_s \frac{\partial \tilde{T}_s}{\partial t} = \frac{\partial}{\partial \tilde{x}}\left(\kappa_s \frac{\partial \tilde{T}_s}{\partial \tilde{x}}\right) - \frac{\kappa}{ws}\frac{\partial \tilde{T}}{\partial \tilde{y}}\bigg|_{\Gamma_{\text{plate}}} \tag{A20}$$

when terms of $O(\eta^2)$ are ignored.

[1] A. A. Atchley, H. E. Bass, and T. J. Hofler, "Development of nonlinear waves in a thermoacoustic prime mover," in *Frontiers of Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Elsevier, London, 1990), pp. 603–608.

[2] G. W. Swift, "Analysis and performance of a large thermoacoustic engine," J. Acoust. Soc. Am. **92**, 1551–1563 (1992).

[3] S. Backhaus and G. W. Swift, "A thermoacoustic-Stirling heat engine," J. Acoust. Soc. Am. **107**, 3148–3166 (2000).

[4] M. E. Poese and S. L. Garrett, "Performance measurements on a thermoacoustic refrigerator driven at high amplitudes," J. Acoust. Soc. Am. **107**, 2480–2486 (2000).

[5] S. Karpov and A. Prosperetti, "Nonlinear saturation of the thermoacoustic instability," J. Acoust. Soc. Am. **107**, 3130–3147 (2000).

[6] V. Gusev, "Asymptotic theory of nonlinear acoustic waves in a thermoacoustic prime-mover," Acustica **86**, 25–38 (2000).

[7] N. Cao, J. R. Olson, G. W. Swift, and S. Chen, "Energy flux density in a thermoacoustic couple," J. Acoust. Soc. Am. **99**, 3456–3464 (1996).

[8] A. S. Worlikar and O. M. Knio, "Numerical simulation of a thermoacoustic refrigerator. I. Unsteady adiabatic flow around the stack," J. Comput. Phys. **127**, 424–451 (1996).

[9] A. S. Worlikar, O. M. Knio, and R. Klein, "Numerical simulation of a thermoacoustic refrigerator. II. Stratified flow around the stack," J. Comput. Phys. **144**, 299–324 (1998).

[10] N. Rott, "Damped and thermally driven acoustic oscillations in wide and narrow tubes," Z. Angew. Math. Phys. **20**, 230–243 (1969).

[11] G. W. Swift, "Thermoacoustic engines," J. Acoust. Soc. Am. **84**, 1145–1180 (1985).

[12] M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Nonlinear two-dimensional model for a thermoacoustic engine," 1st International Workshop on Thermoacoustics, 's-Hertogenbosch, The Netherlands, 23–25 April 2001.

[13] C. Lawrenson, B. Lipkens, T. S. Lucas, D. K. Perkins, and T. W. Van Doren, "Measurements of macrosonic standing waves in oscillating closed cavities," J. Acoust. Soc. Am. **104**, 623–636 (1998).

[14] B.-T. Chu and R. E. Apfel, "Acoustic radiation pressure produced by a beam of sound," J. Acoust. Soc. Am. **72**, 1673–1687 (1982).

[15] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon, New York, 1987).

[16] W. F. Ames, *Numerical Methods for Partial Differential Equations*, 3rd ed. (Academic, San Diego, 1992).

[17] R. L. Panton, *Incompressible Flow*, 2nd ed. (Wiley, New York, 1996), Sec. 16.4.

[18] D. T. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000), Chap. 7.

[19] A. A. Atchley, "Analysis of the buildup of oscillations in a thermoacoustic prime mover," J. Acoust. Soc. Am. **95**, 1661–1664 (1994).

[20] M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Linear and nonlinear frequency shifts in acoustical resonators with varying cross sections," J. Acoust. Soc. Am. **110**, 109–119 (2001).

[21] M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Acoustic streaming generated by standing waves in narrow channels," J. Acoust. Soc. Am. **110**, 2653(A) (2001).

[22] J. Matthews and R. L. Walker, *Mathematical Methods of Physics*, 2nd ed. (Addison-Wesley, New York, 1970), Eq. (15-69).

Investigations of sound waves generated by the Hall effect in electrolytes

Angelo J. Campanella[a)]
3201 Ridgewood Drive, Columbus (Hilliard), Ohio 43026

(Received 30 May 2001; accepted for publication 22 January 2002)

Electrolyte ions moving in a magnetic field produce a velocity component transverse to the electric and magnetic fields. This phenomenon can be alternatively termed as the Hall effect, a magnetohydrodynamic transducer, or a liquid-state electromagnetic transducer. In electrolytes, the viscous drag of the electrolyte fluid is dominant, so the ion deflection soon reaches a terminal velocity dependent only on the electric field strength. The alternating velocity component transverse to the electric and magnetic fields produces a pressure wave due to the viscous drag of the fluid on the ions. Acoustic waves of very high frequencies can be produced since the mass of an ion is very small. Theory for this method of generating sound waves by ultrahigh frequency (UHF) electromagnetic wave impinging on a conducting medium in a magnetic field, causing a conduction current to flow near the surface (skin effect), is presented. Experiments were performed in an electrolyte at frequencies from 8 kHz to 1 MHz to study the effects of electrode size, ionic mobility, and ionic concentration. In the reciprocal effect, ions moving en masse with the particle velocity of a passing acoustic wave induce an alternating voltage between a pair of electrodes in a transverse magnetic field. [DOI: 10.1121/1.1460922]

PACS numbers: 43.38.Dv, 43.35.Rw, 43.30.Yj [SLE]

GLOSSARY OF SYMBOLS

(Vector quantities are listed before their component magnitudes). The unit of dimensionless quantities is listed as ''nondim.''.

Quantity symbol	Quantity	Unit symbol
a_i	traveling wave amplitudes	nondim.
c	velocity of sound in electrolytes	m/s
$\mathbf{f}$	vector force on an ion	N
f_e	electric force amplitude on an ion	N
f_z	Hall or magnetohydrodynamic force amplitude on an ion	N
j	square root of minus one	unity
$k=\omega/c=2\pi/\lambda$	angular wave number	rad/m
m	ion mass	kg
p	acoustic pressure	Pa
q	ion charge	C
s, w, h	wetted electrode length, width, height	m
u	ion velocity vector	m/s
u_x	ion velocity magnitude in x-direction	m/s
u_z	ion velocity magnitude in z-direction	m/s
v_o	received sound wave particle velocity	m/s
$\times$	vector cross-product operator	nondim.
x, y, z	positions in fixed Cartesian coordinates	m
$x_{,t}; y_{,t}; z_{,t}$	partial first derivatives w.r.t. time	m/s
$x_{,tt}; y_{,tt}; z_{,tt}$	partial second derivatives w.r.t. time	m/s^2
Z_O	electromagnetic impedance of free space	$\Omega\cdot$m
i, j, k	unit vectors in x-, y-, and z-directions	unity
α	viscous force constant	N·s/m
β	fluid bulk modulus	Pa
$\delta=(\omega\sigma\mu_1/2)^{-1/2}$	skin depth of a conductor	m
ϵ_1	permittivity of the conducting medium	F/m
$\zeta=J\times B$	Hall force vector	Pa/m
ζ_o	Hall force amplitude	Pa/m
λ	acoustic wavelength	m
μ_1	permeability of the conducting medium	H/m
ξ	particle displacement in z-direction	m
$\xi_{,t}$	acoustic particle velocity	m/s
$\xi_{,tt}$	fluid parcel particle acceleration	m/s^2
$\xi_{,z}$	displacement gradient in z-direction	nondim.
ρ	fluid density	kg/m^3
ρ_c	induced electric charge density	C/m^3
σ	medium electrical conductivity	S/m
ω	angular frequency	rad/s
$A=\alpha/m$	viscous force/ion mass (ratio)	s^{-1}
A	source wave amplitudes	nondim.

a)Electronic mail: a.campanella@worldnet.att.net

Quantity symbol	Quantity	Unit symbol
B	magnetic flux density vector	T
(note)	1 tesla=1 weber per square meter	T, Wb/m^2
B	magnetic flux density magnitude	T
D	reciprocal reflected wave amplitude	nondim.
E	electric field strength vector	V/m
E_O	electric field strength amplitude	V/m
F	vector force on a parcel of electrolyte	N
G_L (ks)	current sensitivity amplitude	nondim.
G_M (ks)	voltage sensitivity amplitude	nondim.
I	electric current	A
J	electric current density vector	A/m^2
$K=j\zeta_o$	electrolyte sound source	m
$(c/\omega)^2/\beta$		
$K=hBv_oc/(j\omega s)$	electrolyte sound receiver	V·m^2
$P=\rho_e/m$	charge to mass ratio for ions	C/kg
$Q=q/m$	charge to mass ratio	C/kg
$S_L=p/I$	loudspeaker sensitivity	Pa/A
X,Z	x- and z-ion motion amplitudes	m
V	reciprocal effect voltage	V

I. INTRODUCTION

A. Preliminary experiments

Trambarulo and Ackerman[1] detected acoustic waves in a test tube of electrolyte carrying an alternating current in a magnetic field. Initially, a glass tube 25 mm in diameter and 600 mm long filled with potassium chloride solution was fitted with planar electrodes 50 mm long, 19 mm wide, and separated by 13 mm placed in one end. One ampere of ac current from an audio signal generator was passed between these electrodes in the presence of a 0.14-T magnetic field. Acoustic waves up to 100 kHz were easily generated. These were detected with a ceramic probe microphone,[2] insulated for underwater use (hydrophone). On removing the tube from the magnetic field, the sound amplitude dropped to 1/5 to 1/10 the original level.

B. Other investigators

The electromagnetic flow meter (Kolin[3]) operates on the reciprocal principle. The Hall voltage across the electrodes is proportional to the flow of the fluid.[4] Grossman and Charwat[5] measured turbulent velocity fluctuations with this method. Longitudinal magnetohydrodynamic waves[6] relate to this reciprocal principle. Moose and Klaus[7] measured the voltage output of a small cell in a magnetic field excited by an acoustic standing wave. Asisake and Mixson,[8] Schreppler,[9] Moose and Hixson,[10] and later Zoltogorsky[11] reported sound generation in electrolytes in a magnetic field.

C. Thermoacoustic waves

Sound generation by rapid thermal expansion, noted in these original tests, has been further investigated by others in water[9,11] and in air.[12–14]

D. Overview

In this study, transduction theory is developed for electrolyte ions in a magnetic field compelled to move by an alternating electric field. This results in a force and subsequent velocity in the third normal direction for both positive and negative ions. Viscous drag presents an alternating pressure in that same direction. A one-dimensional model quantifies and predicts the subsequent acoustic pressure. The reciprocal effect is also modeled where a sound wave passing between the electrodes evokes a voltage between the electrodes. Scaled experiments with alternating currents from 8 kHz to 1 MHz, and external sound waves at 100 kHz, produced data supporting these theories. The Appendix provides theory for generating sound waves at frequencies up to 1 GHz by the skin effect in a conductor in a magnetic field.

II. TRANSDUCTION THEORY

A. Ion moving in crossed electric and magnetic fields: A one-dimensional model

Assume an electrolyte containing positive and negative charged particles (''ions'') that are forced to move by an electric field, **E**, and where the ion velocity, **u**, is directly proportional to **E** (see Fig. 1). The medium resists ion motion with a force, $\mathbf{f}_e$, proportional to ion velocity

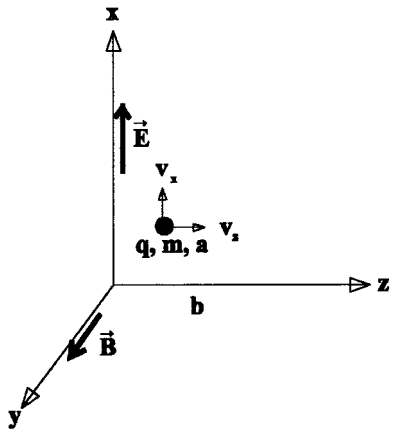


FIG. 1. Ion moving in crossed electric and magnetic fields.

$$\mathbf{f}_e = \alpha \mathbf{u}, \tag{1}$$

where q is the charge on the ion and α is the viscous force constant. The ion velocity, $\mathbf{u}$, is much less than the velocity of light, c', and the velocity of sound, c, therefore, relativistic and nonlinear effects are not expected. The ion force-balance equation for a moving ion of mass, m, by Newton's second law is

$$m(d\mathbf{u}/dt) = q\mathbf{E} - \alpha \mathbf{u}. \tag{2}$$

Now, assume a magnetic field $\mathbf{B}$, normal to the electric field $\mathbf{E}$, throughout the medium. An additional force, $q\mathbf{u}\times\mathbf{B}$, where $\times$ denotes the vector cross product, is exerted on the ion. The new force-balance equation for the moving ion is

$$m(d\mathbf{u}/dt) = q(\mathbf{E} + \mathbf{u}\times\mathbf{B}) - \alpha \mathbf{u}. \tag{3}$$

The electric field is assumed to be along the x axis and the magnetic field to be along the y axis of a Cartesian coordinate system. The vector Hall force in a viscous electrolyte is

$$\mathbf{u}\times\mathbf{B} = -\mathbf{i}u_z + \mathbf{k}u_x. \tag{4}$$

The quantities $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are unit vectors in the x-, y-, and z directions (Fig. 1). In a practical electrolyte, the net x force via the electric field, $\mathbf{E}$, in any fluid parcel is zero since an equal number of positive and negative ions per unit volume are present. However, the z force due to each moving negative ion will be in the same direction as that due to a moving positive ion, since for the latter, both the velocity direction and the force direction will have negative signs.

Let partial differentiation with respect to time be indicated by subscript, i.e., $\partial z/\partial t$ (also u_z) is represented here as $z_{,t}$. Velocity and acceleration components are the time derivatives of ion position (defined here as $x_{,t}$, $x_{,tt}$, $y_{,t}$, $y_{,tt}$, $z_{,t}$, and $z_{,tt}$) so that (3) becomes

$$m(\mathbf{i}x_{,tt} + \mathbf{j}y_{,tt} + \mathbf{k}z_{,u}) = \mathbf{i}(qE - qBz_{,t} - \alpha x_{,t}) - \mathbf{j}\alpha y_{,t} + \mathbf{k}(qBx_{,t} - \alpha z_{,t}). \tag{5}$$

The x-, y-, and z components are equal separately, resulting in

$$mx_{,tt} = qE - qBz_{,t} - \alpha x_{,t}, \tag{6}$$

$$my_{,tt} = \alpha y_{,t}, \tag{7}$$

$$mz_{,tt} = qBx_{,t} - \alpha z_{,t}. \tag{8}$$

Equation (7) allows a solution of the form, $y_{,t} = \mathrm{K}e^{(-\alpha t/m)}$. Motion in the y direction will be damped exponentially with time, which is of no interest here. Let E vary harmonically with time

$$\mathbf{E} = \mathbf{i}E_o e^{(j\omega t)}, \tag{9}$$

where E_o is the magnitude of the electric vector, ω is its angular frequency, and j is the square root of minus 1. Then, x has the form $x = \mathrm{X}e^{(j\omega t)}$, and z the form $z = \mathrm{Z}e^{(j\omega t)}$. Let the amplitudes of the x- and z-ion motions be X and Z. Let $A = \alpha/m$, and $Q = q/m$. From Eq. (8), the z velocity is

$$z_{,t} = Q^2 BE_o e^{(j\omega t)}/[-\omega^2 + Q^2B^2 + A^2 + 2j\omega A]. \tag{10}$$

Equation (10) represents $z_{,t}$ when the frictional term, $\alpha\mathbf{u}$, is linear. Simplification of Eq. (10) is possible if A is dominant. The strongest magnetic induction, $\mathbf{B}$, considered is about 1 T. For a monovalent electrolyte ion, the ionic charge is about 10^5 Coulombs per gram-mole equivalent. A liter of K Cl electrolyte of 1-N concentration contains 1 Faraday (96 494 Coulombs) of either ion. But, for an individual ion, the mass, m, of the K^+ ion is 39 amu plus several water molecules adhered by solvation,[15] a total of perhaps 100 amu. The value of $Q = q/m = 1.6\times10^{-19}/(100\times1.67\times10^{-27}) = 9.6\times10^5$ C/kg for each ion. The value of QB in Eq. (10) is 9.6×10^5 (s^{-1}). Potassium and chloride ions both will have an ionic velocity, $x_{,t}$, of about 6×10^{-8} m/s for an electric field of 1 V/m.[16] The value of $\alpha = qE/x_{,t} = 1.67\times10^{-19}/(6\times10^{-8}) = 2.8\times10^{-12}$ N.s/m. Then, $A = \alpha/m = 2.8\times10^{-12}/(100\times1.66\times10^{-27}) = 1.7\times10^{13}$ s^{-1}. Both A and QB have the dimension of s^{-1}. QB is on the order of 10^6 s^{-1}. Accordingly, the viscous term, A, dominates when $\omega \ll 10$.[13] The frequencies considered here will be less than 10 GHz, where ion movements are quasistatic, and Eq. (10) can be written within 1/2% accuracy as:

$$z_{,t} = (Q/A)^2 BE_o e^{(j\omega t)}. \tag{11}$$

Each ion moving in the z direction exerts a viscous friction force on the medium

$$f_z = \alpha z_{,t} = \alpha(Q/A)^2 BE_o e^{(j\omega t)}. \tag{12}$$

From Eq. (6) and the same dominant terms, the factor $(Q/A)E_o e^{(j\omega t)}$ is also equal to the ion x-velocity amplitude, so that the z-force per ion may be written as

$$f_z = qBx_{,t}. \tag{13}$$

Assume n_+ positive and n_- negative monovalent ions per unit volume in the electrolyte. With reference to Fig. 2, the force F_z exerted on an electrolyte parcel w meters wide, s meters long, and h meters high by ions moving with velocities $x_{+,t} = x_{,t}/2$ and $x_{-,t} = -x_{,t}/2$ is

$$F_z = Bwsh(n_+ x_{,t}.q_+ . + n_- x_{-,t} q_-) = BwshJ = BwshJ_o e^{(j\omega t)}, \tag{14}$$

where $J = (n_+ x_{,t}.q_+ + n_- x_{-,t} q_-)$. This force is uniform across the area, wh, for a uniform current density and magnetic field. It will create an acoustic pressure in the z direc-

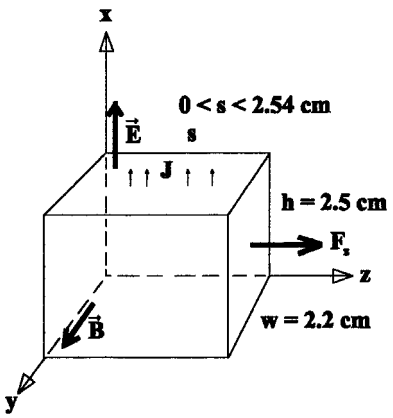


FIG. 2. Force on a fluid electrolyte parcel.

tion (when the electrode length, s, is small compared to an acoustic wavelength) of

$$p_z = BsJ_o e^{(j\omega t)}. \tag{15}$$

Note that the acoustic particle velocity, $\xi_{,t}$, of that wave may differ from the ion velocity, u_z. Thus, if $B=0.1$ T, $s=0.01$ m, $J_o=1000$ A/m^2, then $p_z=1$ Pa.

B. One-dimensional-model of a cell with electrodes long compared to a wavelength

Assume a pair of planar electrodes parallel to the z axis (Fig. 3). These electrodes are both w meters wide in the y direction and s meters long in the z direction. They are h meters apart in the x direction. The space between them contains an electrolyte of conductivity, σ, density, ρ, and bulk modulus, β.

Acoustic pressure is generated uniformly throughout the cell where the current density is $J=J_o e^{j\omega t}$, where J_o is zero everywhere except between $z=0$ and $z=s$. When this region has dimensions comparable to or larger than an acoustic wavelength, the amplitude of the acoustic waves emerging from this region toward $z=+$infinity will depend on the electrode length, s, the magnetic induction B, and the current density J. Acoustic waves originating in the region $0<z<s$ will propagate toward $z=+$infinity without encountering any impedance discontinuity and without divergence for one-dimensional propagation.

Let ζ be the Hall force per unit volume in the z direction produced when the current and magnetic field interact. (Since ζ has dimensions of Pa/m; ζ/ρ has dimensions of acceleration.) For an electrolyte, $\zeta=\mathbf{J}\times\mathbf{B}$. Let ξ be the particle displacement in the z direction and let ζ be the driving term in the acoustic wave equation

$$\rho\xi_{,tt} = \beta\xi_{,zz} + \zeta. \tag{16}$$

Divide by β and replace ρ/β with c^{-2}, where c is the velocity of sound in the electrolyte and the ratio of specific heats is approximately unity. Equation (16) becomes

$$\xi_{,tt}/c^2 = \xi_{,zz} + \zeta/\beta. \tag{17}$$

The driving term ζ can be written with a time factor as $\zeta_o e^{j\omega t}$. A solution of the form $\xi=Ze^{j\omega t}$ is possible where Z contains the conducting cell parameters so that (17) provides

$$Z_{,zz} + \zeta_o/\beta + Z(\omega/c)^2 = 0. \tag{18}$$

Solutions for Z and ξ are of the form

$$Z = a_1 e^{j\omega z/c} + a_2 e^{-j\omega z/c} - K, \tag{19}$$

$$\xi_{\text{in}} = e^{j\omega t}[a_1 e^{j\omega z/c} + a_2 e^{-j\omega z/c} - K], \tag{20}$$

where $K=\zeta_o(c/\omega)^2/\beta$. Outside the cell ($z>s$), ξ takes the form

$$\xi_{\text{out}} = e^{j\omega t}[a_3 e^{j\omega z/c} + a_4 e^{-j\omega z/c}]. \tag{21}$$

Both ξ and its space derivative $\xi_{,z}$ will be continuous across the $z=s$ boundary

$$[\xi_{\text{in}}(z)]_{z=s} = [\xi_{\text{out}}(z)]_{z=s}; \quad [\xi_{\text{in},z}(z)]_{z=s} = [\xi_{\text{out},z}(z)]_{z=s}.$$

If the fluid extends to $+$infinity and there is no impedance discontinuity past for $z>s$, then no wave will return ($a_3=0$). For $z>s$, only outgoing waves exist, having a displacement ξ_{out} of

$$\xi_{\text{out}} = e^{j\omega t}[a_4 e^{-j\omega z/c}]. \tag{22}$$

The amplitude a_4 depends on the boundary conditions at $z=0$ and the length s.

For a rigid backing ($z<0$), ξ is zero there; $[\xi_{\text{in}}(z)]_{z=0}=0$. Within the cell and with the same continuities at $z=s$

$$a_1 = \tfrac{1}{2}Ke^{-j\omega s/c}, \quad a_2 = K[1-\tfrac{1}{2}e^{-j\omega s/c}],$$

resulting in

$$\xi_{\text{in}} = Ke^{j\omega t}[e^{-j\omega z/c} + e^{j\omega s/c}\sin(\omega z/c) - 1]. \tag{23}$$

The acoustic field changes in amplitude and phase in the region between the cell electrodes.

Beyond the electrodes where $z>s$

$$\begin{aligned} a_4 &= K[1-\cos(\omega s/c)]; \\ \xi_{\text{out}} &= Ke^{j\omega(t-z/c)}[1-\cos(\omega s/c)]. \end{aligned} \tag{24}$$

The acoustic pressure is the negative gradient of ξ times ρc^2

$$p(z) = -\rho c^2 \xi_{,z}. \tag{25}$$

Let $\omega/c=k$, where $k/(2\pi)$ is the wave number. It will be shown for (28) from (15) for electrolytes that $K=j\zeta_o/(\beta k^2)$. Note that $\beta=\rho c^2$. From (24) and (25),

$$p_{\text{out}} = \zeta_o e^{j\omega(t-z/c)}[1-\cos(\omega s/c)]/k. \tag{26}$$

The Hall force per unit volume ζ_o is $J_o B$. The current density is $J=I/(sw)$, ζ_o is $I_o B/(sw)$, and (26) becomes

$$p_{\text{out}} = (I_o B/w)e^{j\omega(t-z/c)}[1-\cos(ks)]/(ks). \quad \text{(rigid)} \tag{27}$$

For a given electrode width, w, and when $(ks)\ll 1$, the acoustic pressure increases as the magnetic field and the current increase.

The boundary condition for the free backing is that the acoustic pressure is zero at $z=0$; $[\xi_{\text{in},z}(z)]_{z=0}=0$. Conditions at $z=s$ are unchanged. The displacement and acoustic pressure for $z>s$ are, from (21), (25), $\zeta_o=I_o B/(sw)$, and then (15)

$$\begin{aligned} a_4 &= -K\sin(ks); \quad \xi_{\text{out}} = -Ke^{j\omega(t-z/c)}\sin(ks); \\ \xi_{,z\text{out}} &= -(-jk)Ke^{j\omega(t-z/c)}\sin(ks). \end{aligned}$$

$p_{\text{out}} = -j\beta kKe^{j\omega(t-z/c)}[\sin(ks)] = BsJ_o e^{(j\omega t)}$. For $z\to 0$, $s\to 0$, where $K=BJ_o/(-j\beta k^2)=j\zeta_o/(\beta k^2)$

$$\begin{aligned} p_{\text{out}} &= (\zeta_o/k)e^{j\omega(t-z/c)}[\sin(ks)] \\ &= (I_o B/w)e^{j\omega(t-z/c)}[\sin(ks)]/(ks). \quad \text{(free)} \end{aligned} \tag{28}$$

When the material beyond $z<0$ has the same specific acoustic impedance as that between the electrodes, a wave of amplitude a_3 in (21) will propagate in that direction. And, for $z>s$

$$\begin{aligned} a_4 &= -Ke^{j\omega s/2c}\sin(\omega s/2c) \\ \xi_{\text{out}} &= -Ke^{j\omega(t-z/c+s/2c)}\sin(\omega s/2c). \end{aligned} \tag{29}$$

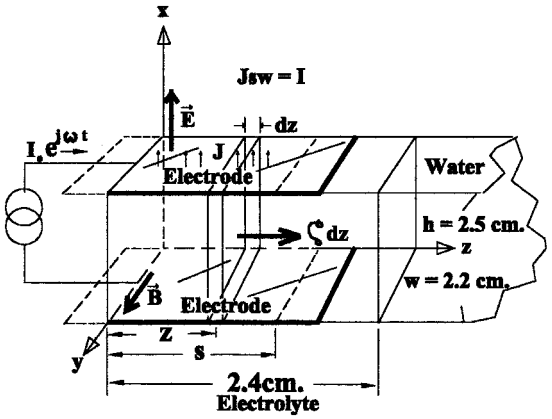


FIG. 3. Electrode cell and the fluid driving (current) force.

$$p_{\text{out}} = \zeta_o e^{j\omega(t-z/c+s/2c)} \sin(ks/2)/k$$
$$= \tfrac{1}{2}(I_o B/w) e^{j\omega(t-z/c+s/2c)}$$
$$\times [\sin(ks/2)]/(ks/2). \quad \text{(matched)} \tag{30}$$

Equations (27), (28), and (30) each have the form $p_z = (I_o B/w) e^{j\omega t} G_L(ks)$, where

$$G_L(ks) = F_i(ks)/(ks). \tag{31}$$

The current sensitivity, S_L, at the cell exit is

$$S_L = p/I = p_z/I = (B/w) G_L(ks). \tag{32}$$

See Fig. 5(a) for a graph of $G_L(ks)$ for a rigid, a free, and a matched backing at $z=0$.

C. The reciprocal effect

Consider an acoustic wave propagating through an electrolyte in the negative z direction and in the presence of both a fixed magnetic field, **B**, in the y direction and a set of fixed electrodes. See Fig. 4. Electrolyte parcels will move with the particle velocity, $\xi_{,t}$, of the acoustic wave, carrying with them the ions thus moving at the same velocity $z_{,t}$, so that $z_{,t} = \xi_{,t}$. By Hall interaction with B, the electric field, E, in Eq. (6) arises from the displacement of positive and negative ions by a distance, x, where $E = \rho_c x/\epsilon_1$. Let $P = \rho_c/\epsilon_1$. As done for Eq. (10), in Eq. (6) let x have the form, $x = Xe^{(j\omega t)}$ and z have the form, $z = Ze^{(j\omega t)} (= \xi)$, but now solving for the ion x displacement resulting from the passage of an acoustic wave ($\xi_{,t} = z_{,t} = j\omega z = p/\rho c$)

$$x = j\omega QBZe^{(j\omega t)}/[\omega^2 - j\omega A + QP]. \tag{33}$$

The value of ρ_c is derived from 96 484 Coulombs per gram-mole, or about 9.6×10^7 Coulombs/m^3 of 1 N KCl electrolyte. The value of Q is 9.6×10^5 C/kg for each ion. The value of ϵ_1 will be that of water, 80 times that of free space; or $8.85\times10^{-12}\times80 = 7.1\times10^{-10}$ F/m, resulting in $P = 1.4\times10^{17}$. A is 1.7×10^{13} (s^{-1}). A very high rollover or relaxation frequency exists according to the denominator of Eq. (33), where ω is of the order of magnitude of A or P; certainly above 10^{10} or 1 GHz, below which the ion movement and electrode voltage can be assumed to be viscous and quasistatic and that $X = Bz_{,t}/P$ and that $E = B\xi_{,t}$. Assume n_+ positive and n_- negative monovalent ions per unit volume being moved with velocity $\xi_{,t} = z_{,t}$. Electrodes of length s, width w, and separation h are placed in the cell (Fig. 4). Hall force displaces positive ions by $X_+ = X/2$ and negative ions by $X_- = -X/2$, causing inducing field $E = (X_+P_+ + X_-P_-) = XP = B\xi_{,t}$. For an acoustic wave particle velocity, $\xi_{,t}$, and a magnetic field strength, **B**, both constant across surface, wh at z, everywhere on that surface the voltage is sensed by the electrodes $V_x(z) = hB\xi_{,t}(z)$. This would also be the case for $s \ll \lambda$. If the cell length, s, were so long as to be comparable to or greater than the acoustic wavelength, λ, the voltage average, V, between the electrodes will be determined by the average of $\xi_{,t}(z)$ across area, sw, from $z=0$ through $z=s$

$$V = (hB/s) \int_0^s \xi_{,t}(z)\, dz. \tag{34}$$

The acoustic particle velocity may have the form

$$\xi_{,t} = \xi_{,to}[e^{j\omega(t+z/c)} + De^{j\omega(t-z/c)}], \tag{35}$$

where D is the relative amplitude of the wave reflected at $z=0$. Integration of (34) yields

$$V = Ke^{j\omega t}[(e^{j\omega s/c} - 1) - D(e^{-j\omega s/c} - 1)], \tag{36}$$

where $K = hB\xi_{,to}c/(j\omega s)$. Note that this cell acts as a particle velocity sensor. The voltage appearing across the electrodes is given by (36), which has the form

$$V = hB\xi_{,to} e^{j\omega t} G_M(ks),$$

the time average of which is

$$V_{\text{rms}} = hB\xi_{,\text{trms}} G_M(ks).$$

The velocity-microphone sensitivity of the cell is

$$M_x = V_{\text{rms}}/\xi_{\text{trms}} = hBG_M(ks). \tag{37}$$

If a reflection by a barrier occurs at $z=0$, then $u_{z=0} = 0$, $D = -1$, and the observed voltage is

$$V = 2KhB\xi_{,to} e^{j\omega t}[\cos(ks) - 1]$$
$$= 2hB\xi_{,to}[\cos(ks) - 1]/(ks). \quad \text{(rigid)} \tag{38}$$

If the surface to the left is free, $D = +1$, and V is

$$V_{\text{rms}}/\xi_{\text{trms}} = 2hB \sin(ks)/(ks). \quad \text{(free)} \tag{39}$$

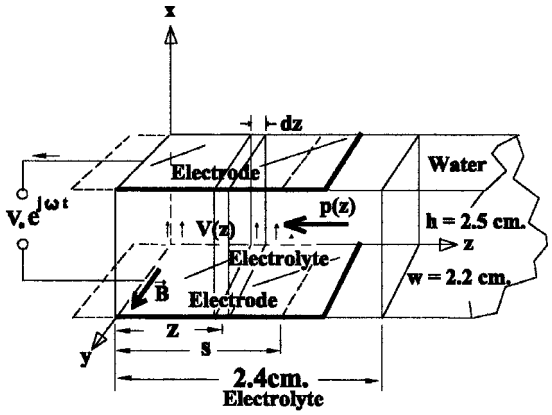


FIG. 4. Electrode cell and the reciprocal (voltage) effect.

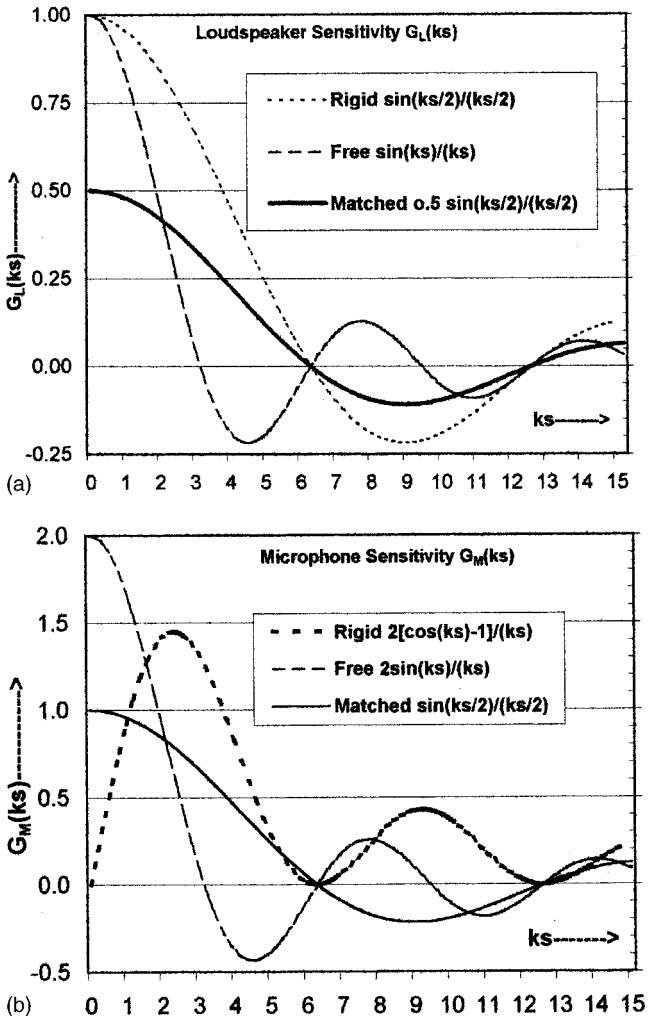


FIG. 5. (a) Current sensitivity function, $G_L(ks)$ for rigid, free, and matched backing conditions. (b) Voltage sensitivity function, $G_M(ks)$ for rigid, free, and matched backing conditions.

If no reflection occurs at $z=0$ (matched fluid impedance), then $D=0$ and V is

$$V_{\rm rms}/\xi_{\rm trms}=hBe^{j\omega s/(2c)}\sin(ks/2)/(ks/2). \quad \text{(matched)} \tag{40}$$

The factor $e^{j\omega s/(2c)}$ may be ignored since the amplitude of an exponential with a purely imaginary argument is always unity. For short electrodes or at low frequencies, $\omega s/c \ll 1$, (38) vanishes, (39) becomes $V_{\rm rms}/\xi_{\rm trms}=2hB$, and (40) becomes $V_{\rm rms}/\xi_{\rm trms}=hB$. The acoustic pressure is $p_{\rm rms}=\rho c\xi_{\rm trms}$, so that $V_{\rm rms}/p_{\rm rms}=hB/(\rho c)$. For greater values of s

$$V_{\rm rms}/p_{\rm rms}=[hB/(\rho c)]G_M(ks) \quad \text{V/Pa}. \tag{41}$$

See Fig. 5(b) for graphic representation of $G_M(ks)$ for rigid, free, and matched backing at $z=0$. When $ks \gg 1$, in all cases the electrodes are generally ''too large,'' producing a small acoustic signal voltage despite electrode extent.

III. MEASUREMENT APPARATUS FOR 8-kHz 1.05-MHz SOUND WAVES IN ELECTROLYTES

A. Basic sound source cell

Electrodes were placed in the center of an open rectangular trough 25 mm high, 36 mm wide, and 900 mm long, located in a magnetic field of 0.14 T. Pulses of alternating current from a 1000-W audio amplifier when passed between electrodes produced an acoustic wave traveling in both directions from the electrodes. Standing waves of sound were found throughout the liquid volume, but no meaningful measurements could be made due to surface reflections. The column of electrolyte was next enclosed in a 2.4-m-long polystyrene tube that also comprised a water-filled delay to avoid electrical cross talk and allowed visual inspection for the undesired air bubbles. Copper-plate electrodes were affixed to a Lucite plug fitted into the end of the tube and sealed with an O-ring. The electrode cell was partitioned from the water-filled tube with a thin polyethylene film. A pulse repetition rate of 5 pps avoided bubble formation on the electrodes from heating.

For this 25-mm-diameter polystyrene tube, uniformity across the tube cross section was assured for frequencies below 23 kHz. For higher modes, pressure wave amplitude was reduced by propagation attenuation, typically 2 dB at 50 kHz. At some test frequencies, structure-borne sound was conducted through the tube wall to the microphone. Sound measurements were possible at frequencies where tube attenuation and harmonic levels were minimal. At higher frequencies, typically between 40 and 80 kHz, measurements were reported only where the dominant signal was due to water-conducted sound traveling as a higher-order mode.

B. Current generation

The electronic system to produce currents from 8 to 100 kHz comprised a gated sine wave fed to a power amplifier providing alternating current to the cell electrodes. The current through the cell electrodes was determined as the voltage across a calibrated 1-ohm series resistor (insert voltage).

C. Electronic measurement system

A Benham precision narrow-band heterodyne wave analyzer was used to isolate the converted waveform. This arrangement verified that only the fundamental was affected by the magnetic field. Direct observation of the intermediate frequency (IF) 450-kHz waveform voltage provided the most reliable and precise acoustic pressure wave amplitude measurement. The peak value of this waveform was measured by a peak-reading vacuum tube voltmeter (VTVM). A precision step attenuator in the signal chain was adjusted for each datum to produce a predetermined output voltage on the VTVM display for all test data. The hydrophone signal measurement system was calibrated daily with an insertion voltage in series with the hydrophone ground. The peak-reading VTVM measured all signals, with insert voltages having the same time impulse format as the acoustic signal.

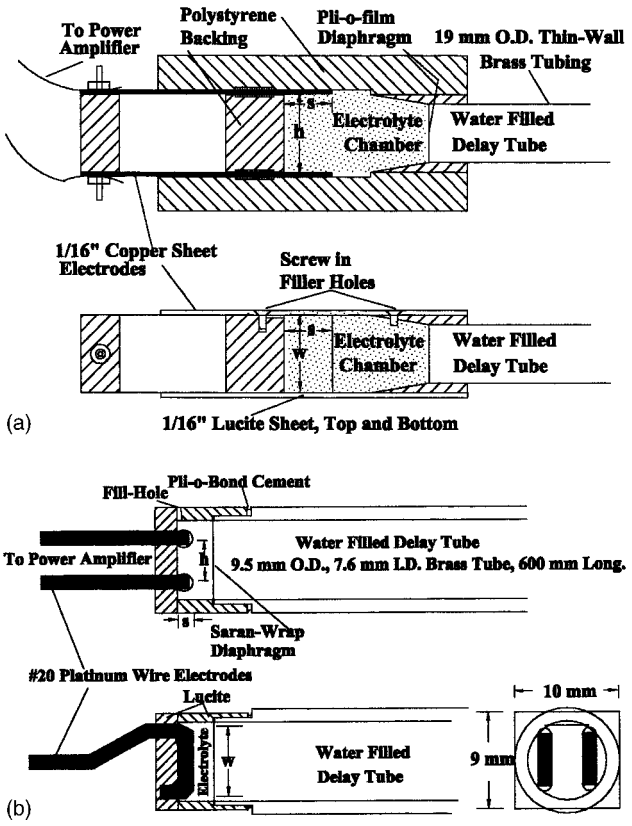


FIG. 6. (a) Electrolyte cell with variable length electrodes and a delay tube. This was used for the low-frequency measurements. (b) High-frequency electrolyte cell and delay tube.

D. Thermoacoustic anomaly

Sound waves at twice the fundamental frequency were observed in 1954–1955. This harmonic wave amplitude varied as the square of the current density. This harmonic wave amplitude was minimized by using electrode spacing, *w*, as wide as possible. In 1986, Shreppler[9] showed that this was a thermoacoustic pressure wave, being instantaneous thermal expansion from heating during each half-wave cycle of alternating current, in agreement with the current density-squared observation.

IV. QUANTITATIVE MEASUREMENT RESULTS

A. Low-frequency measurements

A rectangular cell was constructed of polystyrene blocks and sheet polymethylmethacrylate (''Lucite''). Sliding electrodes spaced 25 mm apart were placed in it [see Fig. 6(a)]. This cell was fitted into the end of a 19-mm-diameter, 2.7-m-long water-filled brass tube (0.75-mm wall thickness). The cell top was covered with a Lucite sheet. The water-filled tube was separated from the cell electrolyte with a polyethylene film dam. The joints were sealed with contact cement. Electrodes were of 1.5-mm sheet copper that could be slid in and out of the electrolyte-filled cell to vary the dimension, *s*, from 0 to 24 mm. The brass delay tube was filled with distilled water previously degassed by boiling. The smaller 19-mm-diameter assured uniformity across the tube cross section for frequencies below 30 kHz.

TABLE I. Measured relative electrolyte p/I values.

Electrolyte	Concentration normal	Frequency, kHz	p/I dB *re* KCl
KCl	1.0	80	0
$ZnSO_4$	1.0	80	1.1
$(NH_4)_2SO_4$	6.0	80.7	−2.5

During initial tests at frequencies of 8 and 14 kHz, the sound emitted by the cell diminished due to electrolyte heating, exceeding the small variation expected according to $\sin(ks)/(ks)$. This diminishment was caused by the reduction of cell acoustic impedance at the cell walls by the gas bubbles forming on the electrodes. Thereafter, the alternating current was pulsed, and only when measurements were being made. The data at 66, 85, and 100 kHz showed the maximum value near $s/\lambda = 0$ did not occur because of an electrode end effect. The equivalent added length, s_o, was estimated to be at lest 5 mm by extrapolating the relative conductance of the cell vs *s* to a zero conductance. Curve fitting of the variation of p/I vs *s* indicated 9 mm to be a better estimate of s_o.

The ratio of the acoustic pressure amplitude at the hydrophone to the current through the cell versus electrolyte concentration, electrode length, *s*, current density, *J*, ionic mobility, and frequency was evaluated. Table I shows p/I determined for three electrolytes. Measurement error was ±1 dB. Over the range of electrolyte concentration from 0.03 to 3 normal, the value of p/I was found to decrease slightly with increasing concentration due to the change of specific acoustic impedance and the lessened degree of dissociation of the higher-concentration solutions.

The acoustic impedance mismatch between electrolyte and water causes the p/I ratio relative to that for KCl to decrease for the more dense solutions. Electrolytes of 3-normal and 6-normal ammonium sulfate were tested for high current density linearity. The variation of p/I with s/λ was recorded at frequencies of 8, 14, 66, 85, and 100 kHz. For *J* less than 1.5×10^5 amperes per square meter the value of p/I is constant. At higher current densities, the amplifier produced clipped and distorted sine waves, though the acoustic pressure could be linear with current density.

A quantitative check of cell sound output vs electrode length, *s*, using Eq. (30), was performed at a frequency of 84.5 kHz. The inserted electrode length was varied as $0<s<25$ mm. The magnetic field *B* was 0.14 T. The electrolyte was a 1.0 normal potassium chloride solution having a sound velocity estimated to be 1550 m/s. The current, *I*, is *Jsw*. The asymptotic current sensitivity $(p/I)_a$ is given by Eq. (32) when $ks \ll 1$, as $|p_{ozout}| = s\zeta_o = sJB = IB/w$, or $(p/I)_a = B/w$.

To evaluate $(p/I)_a$ in SI units, the magnetic field is expressed in tesla, T, the current is expressed in common amperes, and the acoustic pressure is expressed in pascals. Accordingly, the expected value of $(p/I)_a = 6.4$ pascals/ampere. Figure 7 shows the measured underwater sound field in pascals per ampere vs the effective electrode length/wavelength, s/λ. When $s=0$, there remains a ''virtual length.'' Experimental data where the current that flows when the electrodes just enter the electrolyte indicates a virtual length, s_o, of about 9 mm. Theoretical values of p/I are given by Eq. (27)

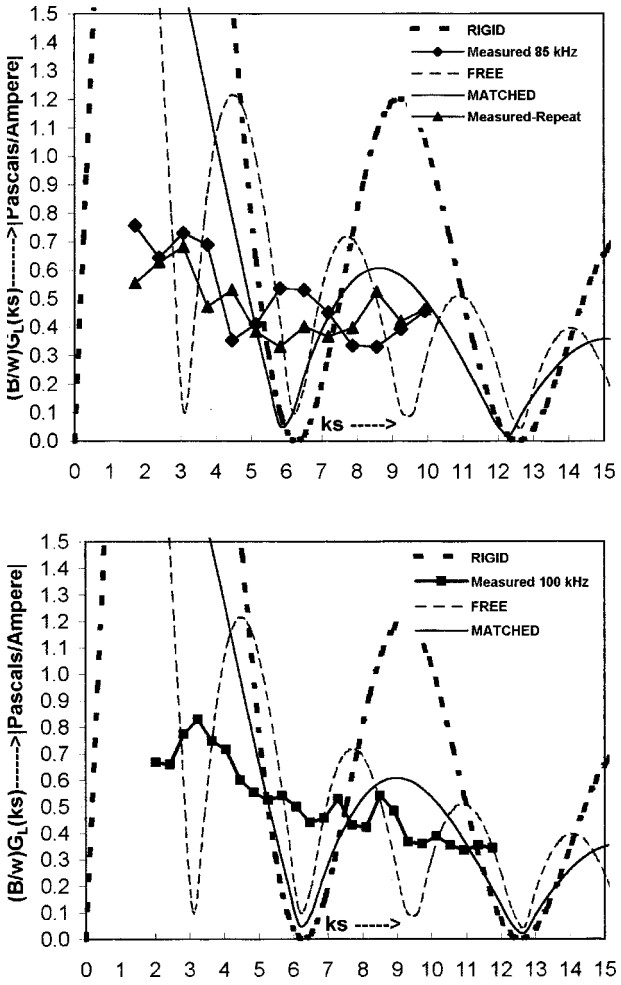


FIG. 7. (a) Measured current sensitivity of the cell at 84.5 kHz. $s_o = 5$ mm. (b) Measured current sensitivity of the cell at 100 kHz. $s_o = 5$ mm.

(rigid), (28) (free), and (29) (matched) backings. There remains consideration of the averaging within the electrode length. The $G_L(ks)$ function for the free-cell-end condition is only about 0.1 in the range of s/λ tested. Values of p/I on the order of 0.5 pascals/ampere are seen to occur, in general agreement with the predicted values for the rigid or the matched backing.

B. High-frequency measurement

A power amplifier tuning from 150 kHz to 1.05 MHz with a maximum power output of 700 watts was used to drive a smaller cell. The electrodes were 20 gage ($s = 0.81$ mm) platinum wires 5.5 mm wide (w), 3.1 mm apart (h), and set flush to one end of a small Lucite cell [see Fig. 6(b)]. A Saran-wrap window, 6 micrometers thick, separated the electrolyte from the delay tube water. A 600-mm-long water-filled brass tube of 9.5-mm outer diameter, 7.6-mm inner diameter, was used as delay tube. A radar magnetron magnet presented a 0.75-T magnetic field in a 9.5-mm gap. A short ceramic probe hydrophone[2] was used to detect acoustic waves produced. Acoustic waves up to 1 MHz were generated and detected. The sound attenuation through the delay tube was significant at most frequencies except 360 kHz and

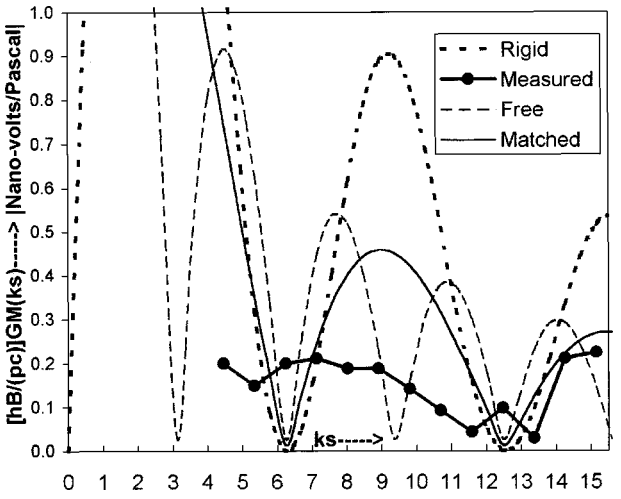


FIG. 8. Measured voltage sensitivity of the cell at 110 kHz. $s_o = 10$ mm.

1.05 MHz. Attenuation was found to be less than 3 dB at 1.05 MHz. The alternating current pulse was as short as 50 microseconds.

The theoretical value $(p/I)_a = B/w$ was 0.75/.0055 $= 136$ pascals per ampere. When a 1.05-MHz ($\lambda = 1.5$ mm) alternating current of 6.7 amperes was passed through the cell, the sound level in the transfer tube was found to be 158 pascals, producing a (p/I) value of 24 pascals/per ampere, implying a value of $G_L(ks) = 0.18$. The electrode length, s, was 0.81 mm, while s_o was previously found to be about one-fourth the spacing, h, or 0.78 mm, providing a value of $s + s_o = 1.6$ mm. These provide a value of ks of about 7. Figure 5(a) shows the matched backing $\frac{1}{2}\sin(ks/2)/(ks/2)$ value to be about 0.05. The rigid $[\cos(ks)-1]/(ks)$ value and the free $\sin(ks)/(ks)$ value each are about 0.1, favoring either as a fair agreement.

C. Reciprocal effect measurement

The low-frequency cell was tested as a velocity hydrophone. The ac voltage appearing across the electrodes was measured when an acoustic wave passed in the z-direction between the electrodes.

The source of the acoustic waves was a Brush-type *BM* 102*A* piezoelectric transducer mounted at one end of a water-filled delay tube. This was coupled to the delay tube with a water-filled conical brass coupler 125 mm long. The acoustic pressure was also monitored along the axis of the electrode chamber with a ceramic probe hydrophone inserted through the backing block, with a voltage sensitivity of minus 122.5 dB $re{:}1$ V/Pa, calibrated in a water tank.[2]

For sound at 110 kHz, the open-circuit voltage was compared to that predicted by Eqs. (38), (39), and (40) in Fig. 8. The sound pressure along the direction of wave propagation in the cell was scanned with the hydrophone to determine that the standing wave ratio was near unity, in turn implying that the backing termination impedance was nearly matched to that of the liquid. Pertinent test values are $h = 25$ mm, $B = 0.14$ T, $\rho = 1000$ kg/m^3, $c = 1{,}550$ m/s, and $V_{\rm rms}/p_{\rm rms} = 2.3\times10^{-9}$ V/pa. The measured (V/p) was about -200 dB $re{:}1$ V/Pa including the factor value of $G_M(ks)$. Figure 8

shows these measured values relative to the predicted (V/p) when s_o is assumed to be 10 mm. Evidence of matched $\sin(ks/2)/(ks/2)$ behavior is seen for ks values from 4 to 15, but with an amplitude of only half that predicted. Klaus,[7] on the other hand, observed a magneto-hydrodynamic potential several times that which he predicted.

V. CONCLUSIONS

The amplitude of the acoustic waves produced by the Hall effect is dependent on the magnetic field strength, the electrode current, and the geometry of the electrode chamber. The impedance of the backing surface is significant, with a free surface being preferred for quantitative measurements of the effect. The end effect of the electrodes amounts to 1/5 to 1/2 the electrode spacing. Other than the impedance change caused by the electrolyte density, sound wave amplitude is independent of the electrolyte concentration. The effect is linear with current density at least up to 1.5×10^5 amperes per square meter within experimental error. At 1 MHz, the effect did not appear to deviate from linearity within the accuracy of the measurements at a current density 1.8×10^5 amperes per square meter.

The generation and the reciprocal effect showed agreement with theory when the spatial average of the sound field between the electrodes is considered.

The generation of second harmonic sound waves by a separate effect, thermal expansion each half-wave cycle, was demonstrated. This effect has been quantified by Mazzola[14] in air and by Schreppler[9] in water. Such rapid thermal expansion is now utilized to generate surface acoustic waves (SAW) in solids.

ACKNOWLEDGMENTS

The Physics Department of the Pennsylvania State University furnished the magnet, the generation and measurement apparatus, and some stipend funding. Appreciation is expressed to Dr. Eugene Ackerman for his guidance throughout these theoretical and experiment efforts. Robert Farwell of ORL (now ARL) calibrated the long hydrophone and William Holak calibrated the short hydrophone.[2] This investigation was performed by the author at the Pennsylvania State University as partial fulfillment of the requirements for the Ph.D degree granted in June, 1955.[17] In the interim, the results of this work were codified into the present form. Recent reviewers provided useful comments and recommendations.

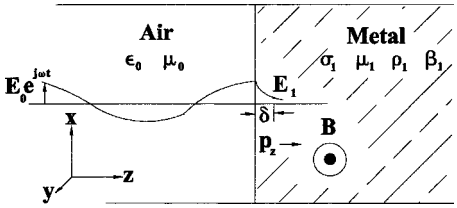


FIG. 9. Electromagnetic wave incident on a conductor in a magnetic field.

APPENDIX: ELECTRODE LENGTH LIMIT BY THE "SKIN EFFECT" AT HIGH FREQUENCIES

Write e^x as $\exp[x]$. At very high frequencies currents in conductors remain near the surface. (See Fig. 9.) Suppose a plane electromagnetic wave of electric intensity $E_e = E_o \exp[j\omega(t-z/c')]$, where c' is the velocity of light, is incident on material of conductivity σ_1, magnetic permeability μ_1, density ρ_1, and bulk modulus β_1, and in the presence of a magnetic field **B** normal to **E**. Most of the electrical energy will be reflected toward-$z>$-infinity. Some electrical energy will enter the conductor when σ_1 is finite. The resulting surface-and near-surface current density, **J**, will produce acoustic waves by interaction with the magnetic field. The electric field within the conductor is[18]

$$E_1 = E_o[(\omega\epsilon_o\mu_1)/(\sigma_1\mu_o)]^{1/2}\exp[j\omega t+(1+j)z/\delta_1]. \tag{A1}$$

The dielectric permittivity of the material where $z<0$ is ϵ_o. This expression is valid when **J** in the conductor is much greater than the displacement current. The current density in the material is $\mathbf{J}=\mathbf{E}_1\sigma_1$. The quantity $\delta_1=(\omega\sigma_1\mu_1/2)^{-1/2}$ is the "skin depth," the distance into the material where the surface current has decreased to $1/e$ of its surface value. Let $Z_o=(\mu_o/\epsilon_o)^{1/2}$, the electrical characteristic impedance of free space. The current density amplitude can be written as

$$J = E_o\{2/(Z_o\delta_1)\}^{1/2}\exp[j\omega t+(1+j)z/\delta_1]. \tag{A2}$$

The driving force ζ per unit volume is $\mathbf{J}\times\mathbf{B}$, with ζ_o

$$\zeta = \mathbf{J}\times\mathbf{B} = \zeta_o\exp[j\omega t+(1+j)z/\delta_1]\mathbf{k}, \tag{A3}$$

$$\zeta_o = J_oB = E_oB\{2/(Z_o\delta_1)\}^{1/2}. \tag{A4}$$

The acoustic wave [Eq. (17)] in this material with z-dependent driving is

$$\zeta_{tt}/{c_1}^2 = \zeta_{,zz}+\zeta/\beta_1. \tag{A5}$$

TABLE II. Metal conductors and skin effect parameters.

Conductor	Conductivity S/m	Sound velocity m/s	Rollover frequency ω_r (MHz)	Skin depth (mm) at ω_r	Acoustic wavelength, mm
Aluminum (Al)	3.54×10^7	6420	292	0.005	0.022
Copper (Cu)	5.80×10^7	5010	291	0.004	0.017
Gold (Au)	4.10×10^7	2030	34	0.014	0.060
Silver (Ag)	6.14×10^7	3650	164	0.005	0.022
Steel ($\mu_1=200\ \mu_0$)	0.50×10^7	5941	35	0.038	0.168
Solder	0.60×10^7	4000±	19	0.047	0.208
Mercury (Hg)	0.104×10^7	1451	0.44	0.746	3.3
Electrolytes	10±	1500±	4.5 Hz	75 meters	330 meters

When no reflections occur for $z>0$, the acoustic particle displacement has the form

$$\xi=\zeta_o\delta_1{}^2\{c_1(1+j)/(j\omega\delta_1)\exp[j\omega(t-z/c_1)] - \exp[j\omega t-z(1+j)/\delta_1]\}/\{\beta_1((\omega\delta_1/c_1)^2+2j)\}. \quad \text{(A6)}$$

The acoustic pressure, p_1, in the conductor is $\beta_1\xi_{,z}$

$$p_1=-\zeta_o\delta_1(1+j)\{-j\exp[j\omega(t-z/c_1)] + \exp[j\omega t-z(1+j)/\delta_1]\}/\{(\omega\delta_1/c_1)^2+2j\}. \quad \text{(A7)}$$

An outgoing wave $\exp[j\omega(t-z/c)]$ and a damped wave, $\exp[j\omega t-z(1+j)/\delta_1]$ are generated. The damped wave is not significant for $z\gg\delta_1$. The acoustic pressure p_z of the remaining outgoing wave, using (A4), is

$$p_z=\{E_oB(1+j)\sqrt{2}\exp[j\omega(t-z/c_1)]\}/\{Z_o((\omega\delta_1/c_1)^2+2j)\}, \quad \text{(A8)}$$

$$|p_z|=2\{E_oB\}/\{Z_o((\omega\delta_1/c_1)^4+4)^{1/2}\} =\{E_oB\}/\{Z_o(\omega^2/(\sigma_1{}^2\mu_1{}^2c_1{}^4)+1)^{1/2}\}. \quad \text{(A9)}$$

When the ω term in the denominator is small compared to unity, the acoustic wavelength is greater than the skin depth and the maximum possible outgoing wave amplitude is essentially independent of frequency. Above the rollover frequency $\omega_r=\sigma_1\mu_1c_1{}^2$ (e.g., 35 MHz for steel, and only 4.5 Hz for sea water), the maximum possible amplitude falls off as $1/\omega$. This is not a limiting frequency; just that above which the maximum possible acoustical output decreases with frequency (see Table II). Above this frequency, the acoustic wavelength is less than the skin depth.

Electrolytes suffer a diminishing output at high acoustic frequencies. Aluminum, copper, and silver could be used to generate sound waves up to at least 100 to 300 MHz. An ultrahigh frequency sound source based on this principle could be used to study relaxation effects in liquids. The sound source would be an electromagnetic wave incident on a metallic ribbon or film immersed in the liquid under study and in a magnetic field. Conventional pulsed radar methods would be used for both transmitter and receiver with a dense metal plate or a free surface to reflect the sound waves back to the ribbon, allowing acoustic wave detection by the reciprocal effect.

[1] Personal communication, Paul Trambulo and Eugene Ackerman (1953).
[2] E. Ackerman and W. Holak, "Ceramic probe microphones," Rev. Sci. Instrum. **25**, 857–861 (1954).
[3] A. Kolin, "An alternating field induction flow meter of high sensitivity," Rev. Sci. Instrum. **16**, 109–116 (1945).
[4] W. G. James, "An induction flowmeter design suitable for radioactive liquids," Rev. Sci. Instrum. **22**, 989–1002 (1951).
[5] L. Grossman and A. Charwat, "The measurement of turbulent velocity fluctuations by the method of electromagnetic induction," Rev. Sci. Instrum. **23**, 741–747 (1952).
[6] N. S. Anderson, "Longitudinal magneto-hydrodynamic waves," J. Acoust. Soc. Am. **25**, 529–532 (1953).
[7] P. H. Moose and R. F. Klaus, "Experimental observation of magneto-acoustic fields," J. Acoust. Soc. Am. **74**, 1066–1068 (1983).
[8] H. Ajisaka and E. L. Hixson, "Transduction principles of a liquid state electromagnetic transducer," J. Acoust. Soc. Am. **77**, 1933–1938 (1985).
[9] S. C. Schreppler, "Magnetohydrodynamic underwater acoustical transducer," ARL, University of Texas, ARL-TR-86-30, 1 December 1986.
[10] H. Ajisaka and E. L. Hixson, "Measurements in a liquid state electromagnetic transducer," J. Acoust. Soc. Am. **81**, 1962–1971 (1987).
[11] B. Zoltogorsky, "The generation of low-frequency acoustic waves in electrolytes," J. Acoust. Soc. Am. **103**, 3075(A) (1998).
[12] W. A. Saxton, "Transducers for exciting and detecting acoustic waves in discharge-type plasmas," J. Acoust. Soc. Am. **37**, 965–973 (1965).
[13] K. J. Nygard and G. Metz, "Generation and detection of acoustic waves in pulsed electrical discharges," J. Acoust. Soc. Am. **44**, 1566–1569 (1968).
[14] M. S. Mazzola and M. Molen, "Modeling of a dc glow plasma loudspeaker," J. Acoust. Soc. Am. **81**, 1972–1978 (1987).
[15] H. J. Creighton, *Electrochemistry, Volume 1, Principles* (Wiley, New York, 1943), p. 40.
[16] Reference 15, p. 133, Table XXXV.
[17] A. J. Campanella, *Generation and Detection of Acoustic Waves by Means of the Hall effect in Electrolytes*, J. Acoust. Soc. Am. **27**, 1005(A) (1955).
[18] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941).

A theoretical study of structural acoustic silencers for hydraulic systems[a)]

Sripriya Ramamoorthy and Karl Grosh[b)]
Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2125

John M. Dodson
Cube 1E-C09, MD X7, Experimental Vehicle Building, 20800 Oakwood Boulevard, Dearborn, Michigan 48124-4076

(Received 19 July 2001; revised 11 January 2002; accepted 8 February 2002)

Theoretical studies show that the introduction of an in-line structural acoustic silencer into a hydraulic system can achieve broadband quieting (i.e., high transmission loss). Strategies for using structural acoustic filters for simultaneously reducing reflection and transmission by tailoring the material properties are studied. A structural acoustic silencer consists of a flexible layer inserted into nominally rigid hydraulic piping. Transmission loss is achieved by two mechanisms—reflection of energy due to an impedance mismatch, and coupling of the incoming acoustic fluctuations to structural vibrations thereby allowing for the extraction of energy through losses in the structure. Structural acoustic finite element simulations are used to determine the transmission loss and evaluate designs. Results based on the interaction of orthotropic and isotropic plates with variable geometry, operating in heavy fluids like water and oil, are presented. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1466866]

PACS numbers: 43.50.Gf, 43.20.Mv, 43.40.Fz [MRS]

I. INTRODUCTION

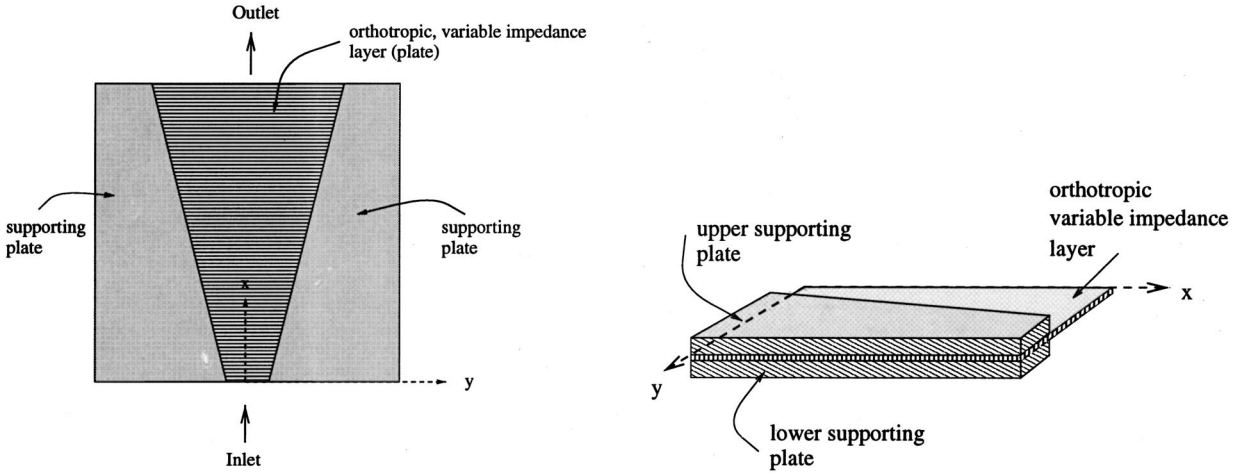
It has long been a goal of the noise control community to develop methods to achieve broadband quieting for waveguide systems. The propagation of acoustic energy in a hydraulic system can be classified into three categories (Martin, 1984): fluid-borne noise which is the result of fluid pressure fluctuations generated by the pump (Edge and Johnston, 1990a, b), structure-borne noise resulting from vibrations within the system's structure, and air-borne noise due to the coupling of the surface vibrations to the surrounding air. Airborne and structure-borne noise are caused to a large extent by fluid-borne noise. Therefore, it is of considerable practical interest to reduce the fluid-borne noise.

Mufflers, quarter wave length resonators, and other reactive devices (e.g., side branches) have dominated quieting designs for the past 80 years (Munjal, 1987; Pierce, 1989). Reactive devices usually provide sound reduction over a narrow band of frequencies around the harmonics of the operating frequency of the device (Dodson *et al.*, 1998; Strunk, 1991). Therefore, this approach is vulnerable to shifts in the operating frequency and variations in the parameters of the system (as they rely on a harmonic cancellation). Insertion of in-line silencers can lead to high static pressure loss and consequently poor engine performance. Flexible metal tuning cable coaxially placed in a section of hose is used to reduce pressure fluctuations in power steering hydraulic circuits. Acoustic performance of hoses is theoretically analyzed by (Munjal and Thawani, 1996) and the experimental evaluation of the noise control performance of tuning cables and hoses is presented in Yu *et al.* (1999) and Dodson *et al.* (1998).

Huang (Huang, 1999; Huang *et al.*, 2000) has performed a theoretical and experimental study of structural acoustic noise control devices. Just as in the current study, a noise control device is placed in the path of flow. A two-dimensional theoretical exploration of the design of a flexible membrane showed the promise of this approach in air (Huang, 1999) and experimental results confirmed some of the predictions (Huang *et al.*, 2000).

In the current study, the design of the mammalian cochlea is adapted for the purposes of devising a noise control device (Ramamoorthy and Grosh, 2000). While the purpose of the biological cochlea is obviously different, its mode of operation exhibits interesting structural acoustic characteristics. For a broad range of excitation frequencies in the cochlea, monofrequency traveling waves propagate until their best places (locations determined by the local impedance) and then are converted to evanescent waves exhibiting rapid spatial decay (Dallos *et al.*, 1996). Such behavior, if attainable, would clearly be desireable. A variable impedance layer (the basilar membrane and organ of Corti in the cochlea) separates the upper and lower ducts of the cochlea (Dallos *et al.*, 1996). The variable impedance layer can be idealized as a series of parallel beams with an increasing width (and thereby decreasing resonance frequency) along the length of the cochlea (i.e., a ''strongly'' orthotropic, variable width plate, as sketched in Fig. 1). In the cochlea, the variable impedance (nearly rigid at the inlet region and compliant at the outlet) gives rise to a space-frequency map, where single frequency excitation excites only a very narrow localized spatial region of the cochlea. It is not the goal of the current work to replicate this sharp localization. Rather, a study of

[a)]Part of this material was presented at the 140th meeting of the Acoustical Society of America, December 2000.
[b)]Author to whom correspondence should be addressed. Electronic mail: grosh@engin.umich.edu



(a) Top view of the variable impedance layer. Variable width of the layer gives rise to a slowly varying impedance.

(b) Perspective view of one–half of the variable impedance layer.

FIG. 1. The orthotropic structural layer is sandwiched between two support plates, and the entire device is enclosed in a housing. Both the housing and the support plates are made from a high-modulus material, nominally considered rigid compared to the flexible variable impedance layer.

the benefits of such a design, using engineering materials rather than biological materials, is undertaken for the purposes of noise control. This study differs from Huang's work (Huang, 1999; Huang *et al.*, 2000) in that plates as well as membranes are used as the structural layer and hybrid modeling techniques for noise control are developed. Also, Huang recently presented two-dimensional theoretical analysis of a membrane of varying compliance interacting with air (Huang, 2001). However, in the current paper, the effectiveness of an orthotropic/isotropic variable impedance plate/membrane for in-line quieting of flow in hydraulic pipes is explored (see Fig. 2) using three-dimensional finite element analysis (FEA). Note that static pressure loading on the structure is not considered in this paper. However, the current design can be made practical by providing static pressure equivalization using a two-duct model as shown in Fig. 3. The presence of the lower cavity allows for symmetric waves (same pressure on both sides of the plate) to propagate, potentially reducing the effectiveness of the device. The design of lower cavity impedance to introduce asymmetry and reject plane waves remains a challenge and will be the focus of future work.

To this end, three-dimensional modeling of wave propagation in fluid-conveying (hydraulic) ducts is performed. The interaction of the acoustic field with orthotropic and isotropic plates with both constant and variable thickness (and/or width) is analyzed. Methods for truncating the finite element computational domain are presented. Transmission loss, power dissipation, and reflectivity are used to compare the theoretical effectiveness of these devices. Results are explained based on wave number analysis. Typical predicted transmission loss for hydraulic systems is on the order of 40 dB.

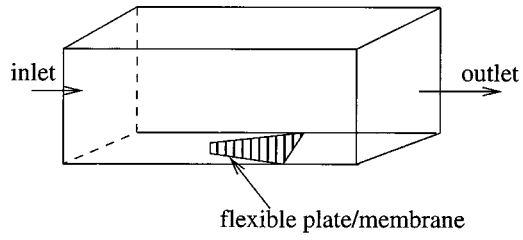


FIG. 2. Structural acoustic silencer.

II. MODELING

The physical configuration of the model is shown in Figs. 4 and 5. In this section, the underlying components of the model are described along with a procedure for numerical approximation.

A. Model geometry

The model consists of a waveguide of rectangular cross section, $L_y \times L_z$, with all walls rigid except one. A part of the wall at $z=0$ consists of a flexible plate or membrane of length L_x and width $b \leqslant L_y$. The fluid structure interface is denoted as $\Gamma = \{0 \leqslant x \leqslant L_x, -b/2 \leqslant y \leqslant b/2, z=0\}$. Here, $b = \epsilon L_y$, where ϵ is the fraction of the duct width occupied by the plate, which will vary as a function of x in the case of a varying width plate. The variation in impedance could also

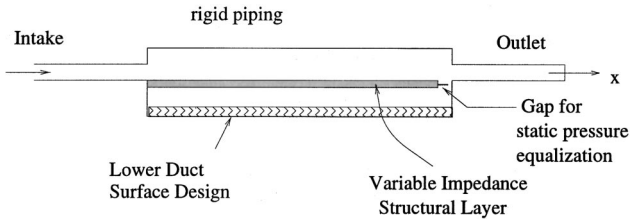


FIG. 3. Schematic of the two-duct model. The upper and lower ducts are separated by a flexible structure and the fluid connected through the gap for static pressure equalization. The lower duct further acts as a backing cavity and may include lossy material at the surface. Only single-duct models are studied in this paper.

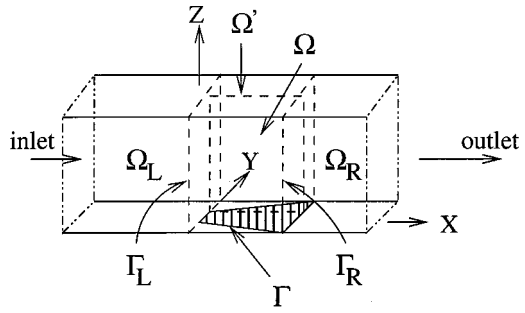


FIG. 4. Duct model showing the computational domains and boundaries. Ω is the 3D computational domain, Ω' is the 2D domain (x-z plane) of finite element discretization for the 2.5D simplified model, Γ is the fluid–structure interface, Γ_L is the left DtN boundary, Γ_R is the right DtN boundary, Ω_L is the infinite region to the left of Γ_L, and Ω_R is the infinite region to the right of Γ_R.

be achieved by spatially varying thickness, or by an appropriate combination of varying thickness and width. The fluid domain is the entire interior of the duct $\{-\infty<x<\infty, -L_y/2\leqslant y\leqslant L_y/2, 0\leqslant z\leqslant L_z\}$, which is broken into three domains: Ω_L to the left of the noise control (or computational) domain Ω, and Ω_R to the right. Here, $\Omega_L=\{-\infty<x<0, -L_y/2\leqslant y\leqslant L_y/2, 0\leqslant z\leqslant L_z\}$ and $\Omega_R=\{L_x<x<\infty, -L_y/2\leqslant y\leqslant L_y/2, 0\leqslant z\leqslant L_z\}$. The finite computational domain is hence $\Omega=\{0\leqslant x\leqslant L_x, -L_y/2\leqslant y\leqslant L_y/2, 0\leqslant z\leqslant L_z\}$.

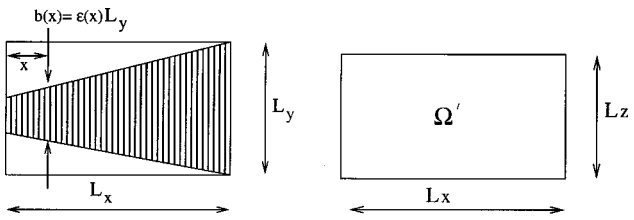
B. Governing equations and assumptions

The region of numerical simulation is the domain Ω including rigid boundaries at $y=-L_y/2$, $y=L_y/2$, $z=L_z$, flexible plate boundary Γ ($z=0$) and Dirichlet-to-Neumann (DtN) boundaries at Γ_L ($x=0$) and Γ_R ($x=L_x$). The DtN boundary is used to convert the infinite domain problem to an equivalent problem on a finite computational domain (described in Sec. II B 2). The fluid is assumed to be compressible and inviscid; fluid viscosity is accounted for heuristically in the plate damping factor. The flexible structure is assumed to be linearly elastic. The input excitation is an incoming right traveling acoustic plane wave impinging on Γ_L. For the frequency range (100–2000 Hz) and geometries of interest, only plane waves propagate in the regions $x<0$ and $x>L$. Time harmonic vibrations of the coupled system are analyzed. A time dependence of $e^{-i\omega t}$ is assumed.

Given these approximations, the coupled fluid–structure governing equations are given by

$$\mathcal{L}_s\mathbf{u}=\hat{\mathbf{n}}p|_\Gamma \quad \text{on } \Gamma \quad \text{structural equations,} \tag{1}$$

$$\nabla^2 p+k^2 p=0 \quad \text{in } \Omega \quad \text{fluid equations,} \tag{2}$$



(a) Bottom view of the model geometry (a varying width plate)

(b) Side view of the model geometry – the 2D discretization domain, Ω'

FIG. 5. Model views to clarify the geometry.

$$\left.\frac{\partial p}{\partial n}\right|_{x=0}=M_L p-2ikp_{\text{in}} \quad \text{on } \Gamma_L \quad \text{left DtN boundary,} \tag{3}$$

$$\left.\frac{\partial p}{\partial n}\right|_{x=L_x}=M_R p \quad \text{on } \Gamma_R \quad \text{right DtN boundary,} \tag{4}$$

$$\frac{\partial p}{\partial n}=\rho_f\omega^2\mathbf{u}\cdot\hat{\mathbf{n}} \quad \text{on } \Gamma \quad \text{Euler's relation,} \tag{5}$$

where $\mathbf{u}=u\hat{e}_z$ is the flexible structure displacement ($\hat{e}_z$ is the unit normal along the z direction), p is the fluid pressure and $\hat{\mathbf{n}}p|_\Gamma$ represents the fluid forcing of the flexible structure, $\hat{\mathbf{n}}$ is the unit outward normal to any bounding surface (on Γ, $\hat{\mathbf{n}}$ points along the negative z direction), ρ_f is the fluid density, $\mathcal{L}_s$ is the plate or membrane operator which acts on the displacement vector, ∇^2 is the Laplacian, $k=\omega/c$ is the acoustic wave number, and M_L and M_R are DtN operators. Equation (5) is Euler's relation that relates the normal component of the acceleration to fluid pressure for an inviscid fluid. The equality of normal component of fluid velocity and velocity of the structure is prescribed on the interface Γ.

For an inhomogeneous orthotropic plate, the structural operator $\mathcal{L}_s$ takes the form

$$\mathcal{L}_s=\frac{\partial^2}{\partial x^2}\left(D_{xx}\frac{\partial^2}{\partial x^2}+D_{xy}\frac{\partial^2}{\partial y^2}\right)+2\frac{\partial^2}{\partial x\,\partial y}\left(D_s\frac{\partial^2}{\partial x\,\partial y}\right)+\frac{\partial^2}{\partial y^2}\left(D_{xy}\frac{\partial^2}{\partial x^2}+D_{yy}\frac{\partial^2}{\partial y^2}\right)-\rho h\omega^2, \tag{6}$$

where (Humphreys and Rosen, 1993)

$$D_{xx}=\frac{E_1E_2}{E_1-\nu_{12}^2E_2}\frac{h^3}{12}, \tag{7}$$

$$D_{yy}=\frac{E_1^2}{E_1-\nu_{12}^2E_2}\frac{h^3}{12}, \tag{8}$$

$$D_{xy}=\frac{\nu_{12}E_1E_2}{E_1-\nu_{12}^2E_2}\frac{h^3}{12}, \tag{9}$$

$$D_s=G_s\frac{h^3}{6}, \tag{10}$$

where E_1 is the complex Young's modulus in the fiber direction (y direction), E_2 is the complex Young's modulus in the direction transverse to the fibers (x direction), G_s is the complex shear modulus, ν_{12} is Poisson's ratio, and h is the thickness of the plate. The terms D_{xx}, D_{xy}, D_{yy}, D_s, h, and b may vary spatially (as a function of x).

A complex Young's modulus is used to include hysteritic damping. Hence, $E_1=E(1-i\eta)$, where E is the Young's modulus, η is the damping factor (viscous effects of the fluid are hueristically included in this factor), and $i=\sqrt{-1}$. Also, $E_2=\varrho E_1$ and $G_s=G\ (1-i\ \eta)$, where ϱ is the orthotropy ratio of the plate material and G is the shear modulus.

The governing structural differential operator for dynamics of a flexible isotropic membrane under uniform tension is given by

$$\mathcal{L}_s = -T\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)-\rho h\omega^2, \tag{11}$$

where ρh is the mass per unit area of the membrane, $T = T_0(1-i\eta)$, where T_0 is the applied tension (per unit membrane width), and η is the damping factor.

1. 2.5-d model

For the rectangular geometry, the cross modes of the duct satisfying rigid wall conditions are cosine functions. These mode shapes may be used to reduce the dimensionality of the problem by decomposing the solution in terms of the duct cross mode number, m. The fluid pressure can be expanded in a modal sum

$$p(x,y,z) = \sum_{m=0}^{M_{\text{modes}}} p_m(x,z)\phi_m(y), \tag{12}$$

where the modes $\phi_m(y) = \cos((m\pi/L_y)(y+L_y/2))$ satisfy the rigid boundary conditions at the walls.

Similarly, the plate displacement $u(x,y)$ can be expanded in a modal sum,

$$u(x,y) = \sum_{j=1}^{n_{\text{modes}}} u_j(x)\psi_j(y), \tag{13}$$

where n_{modes} is the number of cross modes used and $\psi_j(y) = \sin((j\pi/b)(y+b/2))$ represents the jth structural mode in the y direction and satisfies the simply-supported boundary conditions at $|y| = b/2$. Note that ψ_j is also a function of x in the case of a varying width plate, leading to coupled shape-functions in the x and y directions. The structural displacement is zero for the rigid part of the wall at the interface Γ, i.e., for $b/2 < |y| < L_y/2$. The membrane displacement can be expanded in a modal sum similar to the plate displacement [Eq. (13)], though $u_j(x)$ can be a different function for the membrane.

For the reduced model, Eqs. (2) and (5) are modified in terms of the modal pressure p_m and modal plate displacement u_j as

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right)p_m - \beta_m^2 p_m = 0, \quad \text{fluid equations,} \tag{14}$$

$$\frac{\partial p_m}{\partial n} = -\frac{\rho_f\omega^2}{\xi_m}\sum_{j=1}^{n_{\text{modes}}} u_j(x)\chi_{jm} \quad \text{on } \Gamma, \quad \text{Euler's relation,} \tag{15}$$

where $\beta_m^2 = (m\pi/L_y)^2 - (\omega/c)^2$, ω is the forcing frequency, and c is the speed of sound in the fluid (Parthasarathi and Grosh, 2000). The normalization parameters, χ_{jm} and ξ_m, are given by

$$\chi_{jm} = \int_{-b/2}^{b/2} \psi_j(y)\phi_m(y)\,dy, \tag{16}$$

$$\xi_m = \int_{-L_y/2}^{L_y/2} \phi_m^2(y)\,dy. \tag{17}$$

In order to evaluate the computational advantage of the 2.5-d model over the 3-d model, consider the number of flops in each case. The number of flops is approximately $n_{\text{eqns}}(b_w^2+3b_w)$ for the symmetric band solver and $2n_{\text{eqns}}b_w^2$ for the unsymmetric bandsolver, where n_{eqns} is the number of equations being solved and b_w is the bandwidth (Golub and Loan, 1989). The bandwidth of the generalized stiffness matrix is N_zN_y for a 3-d model and N_zM_{modes} for the 2.5-d model, where N_y is the number of pressure nodes along the y direction in a 3-d model (with brick elements), M_{modes} is the number of y modes in the 2.5-d model (with bilinear quadrilateral elements), and N_z is the number of pressure nodes along the z-direction. Also, $n_{\text{eqns}} \approx N_xN_yN_z$ for a 3-d model and $N_xM_{\text{modes}}N_z$ for the 2.5-d model. Therefore, the ratio of computation times for the 2.5-d model and a 3-d model is $t_{2.5-d}/t_{3-d} \approx (M_{\text{modes}}/N_y)^3$. For equivalent model accuracy, N_y will be at least six times M_{modes}. Therefore the 2.5-d model runs nearly 216 times faster than the equivalent 3-d model.

2. DtN boundary condition

In this section, the DtN operators in Eqs. (3) and (4) and their reduced order counterparts are derived. The input plane wave excitation is also incorporated via the DtN boundary condition.

As mentioned in the discussion on the geometry of the model, the fluid domain is broken into three regions. In the domains consisting of rigid walled fluid filled ducts, Ω_L and Ω_R, analytical solutions are known. For the domain Ω, a numerical solution is required due to the presence of the flexible structure. The DtN map and the boundary conditions on Γ_L and Γ_R convert the problem from one on the infinite domain to an equivalent problem on the finite domain, Ω.

The plane wave input pressure is $p_{\text{in}} = \tilde{p}_{\text{in}}e^{i(kx-\omega t)}$, where $k = \omega/c$, and the scattered pressure is denoted by p_{sc}. The total pressure p is given by

$$p = p_{\text{in}} + p_{\text{sc}} \quad (x \leqslant 0). \tag{18}$$

Thus

$$\frac{\partial p}{\partial n} = -ikp_{\text{in}} + \frac{\partial p_{\text{sc}}}{\partial n}, \tag{19}$$

where, for the purposes of the current derivation, n is the outward normal to the boundary at $x = 0$ (i.e., $\partial p/\partial n|_{x=0} = -\partial p/\partial x|_{x=0}$).

Consider general left-going solutions for the scattered pressure:

$$p_{\text{sc}}(x,y,z) = \sum_{m,l} A_{ml}\Phi_{ml}(y,z)e^{-ik_{ml}x} \quad (x \leqslant 0). \tag{20}$$

Here, $\Phi_{ml}(y,z) = \phi_m(y)\zeta_l(z)$, where $\phi_m(y) = \cos((m\pi/L_y)\times(y+L_y/2))$, $\zeta_l(z) = \cos(l\pi z/L_z)$, and $k_{ml}^2 = \omega^2/c^2 - (m\pi/L_y)^2 - (l\pi/L_z)^2$. For k_{ml}, the root with positive real part or imaginary part is chosen for valid left-going solutions. The coefficients of the scattered field are written as a function of $p_{\text{sc}}(0,y,z)$:

$$A_{ml}=\frac{\int_{x=0}\Phi_{ml}(y,z)p_{sc}(0,y,z)\,dydz}{\int_{x=0}\Phi_{ml}^2(y,z)\,dydz}. \tag{21}$$

Since the outward normal to the boundary Γ_L is in the $-x$ direction, $\partial p_{sc}/\partial n|_{x=0}=-\partial p_{sc}/\partial x|_{x=0}$. This gives

$$\left.\frac{\partial p_{sc}}{\partial n}\right|_{x=0}=\sum_{m,l} ik_{ml}\frac{\int_{x=0}\Phi_{ml}(y',z')p_{sc}(0,y',z')\,dy'dz'}{\int_{x=0}\Phi_{ml}^2(y',z')\,dy'dz'} \times\Phi_{ml}(y,z)\equiv M_L p_{sc}, \tag{22}$$

where the left DtN operator, M_L, is given by

$$M_L=\sum_{m,l} ik_{ml}\Phi_{ml}(y,z)\left(\frac{\int_{x=0}\Phi_{ml}(y',z')\,dy'dz'}{\int_{x=0}\Phi_{ml}^2(y',z')\,dy'dz'}\right). \tag{23}$$

For the plane wave pressure input p_{in}, it is possible to show that

$$M_L p_{in}=ikp_{in}. \tag{24}$$

Equations (18), (19), and (22)–(24) are combined to obtain

$$\frac{\partial p}{\partial n}=M_L p-2ikp_{in}. \tag{25}$$

The second term on the right-hand side of Eq. (25) incorporates loading into the model. This DtN condition (or impedance condition) is used to truncate the computational domain at $x=0$.

For the region $x\geqslant L_x$, the total pressure is the transmitted pressure. The general right-going solution is

$$p(x,y,z)=\sum_{m,l} B_{ml}\Phi_{ml}(y,z)e^{ik_{ml}x}, \tag{26}$$

where $\Phi_{ml}(y,z)$ and k_{ml} are the same as defined earlier. The coefficients B_{ml} can be written in terms of the pressure at $x=L_x$:

$$B_{ml}=\frac{\int_{x=L_x}\Phi_{ml}(y,z)p(L_x,y,z)\,dydze^{-ik_{ml}L_x}}{\int_{x=L_x}\Phi_{ml}^2(y,z)\,dydz}. \tag{27}$$

Hence,

$$\left.\frac{\partial p}{\partial n}\right|_{x=L_x} =\sum_{m,l} ik_{ml}\frac{\int_{x=L_x}\Phi_{ml}(y',z')p(L_x,y',z')\,dy'dz'e^{-ik_{ml}L_x}}{\int_{x=L_x}\Phi_{ml}^2(y',z')\,dy'dz'} \times\Phi_{ml}(y,z)e^{ik_{ml}L_x}\equiv M_R p, \tag{28}$$

where n is the outward normal at Γ_R (i.e., $\partial p/\partial n|_{x=L_x}=\partial p/\partial x|_{x=L_x}$). This leads to the expression for right DtN operator, M_R, as

$$M_R=\sum_{m,l} ik_{ml}\Phi_{ml}(y,z)\left(\frac{\int_{x=L_x}\Phi_{ml}(y',z')\,dy'dz'}{\int_{x=L_x}\Phi_{ml}^2(y',z')\,dy'dz'}\right). \tag{29}$$

It is interesting to note that the operator M_L at $x=0$ is the same as the operator M_R at $x=L_x$.

For the 2.5-d formulation, the DtN operators obtained above are modified for the mth mode as

$$\tilde{M}_m=\sum_l ik_{ml}\zeta_l(z)\frac{1}{\alpha_l}\int_{z'=0}^{z'=L_z}\zeta_l(z')\,dz', \tag{30}$$

where $\alpha_l=\int_{z'=0}^{z'=L_z}\zeta_l^2(z')\,dz'$. With this notation, for the mth transverse mode at Γ_L and Γ_R, respectively,

$$\frac{\partial p_m}{\partial n}=\tilde{M}_m p_m-2ik\tilde{p}_{in}\delta_{m0}\frac{L_y}{\xi_m}, \tag{31}$$

$$\frac{\partial p_m}{\partial n}=\tilde{M}_m p_m, \tag{32}$$

where $\delta_{m0}=1$ for $m=0$, $\delta_{m0}=0$ otherwise, and ξ_m is the normalization parameter described in Eq. (17).

C. Variational equations

The variational equations for the given structural model may be obtained via integration of the admissible displacement and pressure weighting functions ($\bar{u}$ and $\bar{p}$, respectively) over the fluid–structure domain (Hughes, 1987). The variational equations may be written as

$$a^s(\bar{u},u)=(\bar{\mathbf{u}},\hat{\mathbf{n}}p)_\Gamma, \tag{33}$$

$$(\nabla\bar{p},\nabla p)_\Omega-(\bar{p}\hat{\mathbf{n}},\rho_f\omega^2\mathbf{u})_\Gamma-(\bar{p},M_L p)_{\Gamma_L}-(\bar{p},M_R p)_{\Gamma_R} -\frac{\omega^2}{c^2}(\bar{p},p)_\Omega=-(\bar{p},2ikp_{in})_{\Gamma_L}. \tag{34}$$

The notation, $(\cdot,\cdot)_\Gamma$, denotes the inner product over any region Γ. The vector displacement $\mathbf{u}$ is $u(x,y)\hat{e}_z$, where $\hat{e}_z$ is the unit vector along the z direction. The outward normal to the flexible structure, $\hat{\mathbf{n}}$, points along the negative z direction. Also, ∇ represents the gradient operator. $a^s(\bar{u},u)$ is the structural operator given by

$$a^s(\bar{u},u)=\int_\Gamma\left(\frac{\partial^2\bar{u}}{\partial x^2}D_{xx}\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2\bar{u}}{\partial x^2}D_{xy}\frac{\partial^2 u}{\partial y^2} +2\frac{\partial^2\bar{u}}{\partial x\partial y}D_s\frac{\partial^2 u}{\partial x\partial y}+\frac{\partial^2\bar{u}}{\partial y^2}D_{xy}\frac{\partial^2 u}{\partial x^2} +\frac{\partial^2\bar{u}}{\partial y^2}D_{yy}\frac{\partial^2 u}{\partial y^2}-\omega^2\rho h\bar{u}u\right)d\Gamma \tag{35}$$

for a plate, and

$$a^s(\bar{u},u)=\int_\Gamma\left(\frac{\partial\bar{u}}{\partial x}T\frac{\partial u}{\partial x}+\frac{\partial\bar{u}}{\partial y}T\frac{\partial u}{\partial y}-\omega^2\rho h\bar{u}u\right)d\Gamma \tag{36}$$

for a membrane. The term on right-hand side of Eq. (34) represents the forcing due to pressure input (plane-wave input is considered in this paper). The 2.5-d reduction of the plate equations is delayed until the next section, where the approximating functions for the fluid pressure and structural displacement in the finite element subspace are introduced.

D. Finite element formulation

A (Bubnov–)Galerkin finite element formulation is obtained from the variational form of the governing equations (Hughes, 1987). Obtaining the finite element discretization using a Galerkin scheme involves the introduction of approximate solutions via the global interpolation functions. Using Eq. (13), the structural displacement u and the corresponding weighting function $\bar{u}$ are approximated as

$$u^h(x,y)=\sum_{j=1}^{n_{\text{modes}}}\sum_{A=1}^{n_{\text{nodes}}} d_{jA}\psi_j(y)N_A^s(x),$$
$$\bar{u}^h(x,y)=\sum_{j=1}^{n_{\text{modes}}}\sum_{B=1}^{n_{\text{nodes}}} c_{jB}\psi_j(y)N_B^s(x), \tag{37}$$

where $N_A^s(x)$ is the interpolation function for the structural displacement (piecewise Hermite functions for the plate and piecewise linear functions for the membrane), the indices A and B range over the finite element nodes, d_{jA} represents the approximate solution at a node of the finite element interpolation for the jth mode, c_{jB} are the weighting function amplitudes, n_{modes} is the number of cross modes used, and n_{nodes} is the number of finite element nodes used along the length of the structure. In this formulation (Bubnov–Galerkin), the weighting functions are same as the approximating functions.

For the 2.5-d model, only the 2-d region (Ω') in the x-z plane is discretized after introducing modal solution in the y direction. The fluid pressure and the corresponding weighting function are approximated as

$$p^h(x,y,z)=\sum_{m=0}^{M_{\text{modes}}}\sum_{C=1}^{M_{\text{nodes}}} p_{mC}\phi_m(y)N_C^f(x,z),$$
$$\bar{p}^h(x,y,z)=\sum_{m=0}^{M_{\text{modes}}}\sum_{D=1}^{M_{\text{nodes}}} p_{mD}\phi_m(y)N_D^f(x,z), \tag{38}$$

where M_{modes} is the number of y-modes used for the fluid pressure, M_{nodes} is the total number of fluid nodes in the 2-d discretization region (Ω'), $N_C^f(x,z)$ is the bilinear shape function used to interpolate the pressure, and indices C and D range over the finite element nodes.

For steady-state dynamics, the matrix equations resulting from the Galerkin formulation (using 2.5-d model for fluid) will have the form

$$\begin{bmatrix}\mathbf{K}^f & \mathbf{Q}^T\\ \mathbf{Q} & \mathbf{K}^s\end{bmatrix}\begin{Bmatrix}\mathbf{p}\\ \mathbf{d}\end{Bmatrix}=\begin{Bmatrix}\mathbf{f}\\ \mathbf{0}\end{Bmatrix}. \tag{39}$$

The acoustic mode shapes uncouple the fluid equations. Assuming the structural mode shapes also uncouple the structural equations (which is the case for a constant width plate using sine functions for transverse mode shapes, whereas for a varying width plate, orthogonal functions are carefully chosen for the transverse mode shapes), the fluid and structural dynamic stiffness matrices, $\mathbf{K}^f$ and $\mathbf{K}^s$, are given by

$$\mathbf{K}^f=\begin{bmatrix}\mathbf{K}_0^f & 0 & \cdots & 0\\ 0 & \mathbf{K}_1^f & \cdots & 0\\ 0 & 0 & \ddots & 0\\ 0 & 0 & \cdots & \mathbf{K}_{M_{\text{modes}}}^f\end{bmatrix},$$
$$\mathbf{K}^s=\begin{bmatrix}\mathbf{K}_1^s & 0 & \cdots & 0\\ 0 & \mathbf{K}_2^s & \cdots & 0\\ 0 & 0 & \ddots & 0\\ 0 & 0 & \cdots & \mathbf{K}_{n_{\text{modes}}}^s\end{bmatrix}. \tag{40}$$

Here, $\mathbf{K}_m^f$ is the matrix corresponding to the mth fluid mode, assembled over all the fluid nodes. The (C,D)th element of the matrix $\mathbf{K}_m^f$ is given by

$$[\mathbf{K}_m^f]_{CD}=\frac{\xi_m}{\rho_f\omega^2}\left\{\int_{\Omega'}(\nabla N_C^f\cdot\nabla N_D^f+\beta_m^2 N_C^f N_D^f)\,d\Omega'-\int_{\Gamma_L}N_C^f(0,z)\tilde{M}_m N_D^f(0,z)\,dz-\int_{\Gamma_R}N_C^f(L_x,z)\tilde{M}_m N_D^f(L_x,z)\,dz\right\}. \tag{41}$$

The structural stiffness matrix is obtained by substituting Eq. (37) into the structural operator, which is defined by Eq. (35) for a plate, and Eq. (36) for a membrane. In the case of constant width (thickness could be either constant or varying), the structural stiffness matrix, $\mathbf{K}^s$, can be simplified. In this case, the (A,B)th element of the jth mode structural stiffness matrix, $\mathbf{K}_j^s$ for a plate, is given by

$$[\mathbf{K}_j^s]_{AB}=\gamma_j\left\{\int_\Gamma D_{xx}N_{A,xx}^s N_{B,xx}^s\,dx+\int_\Gamma\left(D_{yy}\left(\frac{j\pi}{b}\right)^4-\rho h\omega^2\right)N_A^s N_B^s\,dx-\int_\Gamma\left(j\frac{\pi}{b}\right)^2 D_{xy}(N_{A,xx}^s N_B^s+N_{B,xx}^s N_A^s)\,dx+\int_\Gamma 2\left(j\frac{\pi}{b}\right)^2 D_s N_{A,x}^s N_{B,x}^s\,dx\right\}, \tag{42}$$

where $\gamma_j=\int_{-b/2}^{b/2}\psi_j^2\,dy$ and the inferior comma represents partial differentiation. Here, N_A^s and N_B^s are piecewise Hermite interpolation functions in x. For a membrane, the (A,B)th element of the matrix $\mathbf{K}_j^s$ is given by

$$[\mathbf{K}_j^s]_{AB}=\gamma_j\left\{\int_\Gamma TN_{A,x}^s N_{B,x}^s\,dx+\int_\Gamma\left(T\left(\frac{j\pi}{b}\right)^2-\rho h\omega^2\right)N_A^s N_B^s\,dx\right\}. \tag{43}$$

Here, N_A^s and N_B^s are piecewise linear functions in x. For a varying width plate or membrane, the expressions are more involved and their evaluation is facilitated using symbolic manipulators. See Dodson (2001): Dodson and Grosh (2002) for a discussion of the stiffness matrix for a varying width plate.

The fluid–structure coupling matrix $\mathbf{Q}$ is given by

$$\mathbf{Q}=\begin{bmatrix} \mathbf{Q}_{1,0} & \cdots & \mathbf{Q}_{1,m} & \cdots & \mathbf{Q}_{1,M_{\text{modes}}} \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ \mathbf{Q}_{j,0} & \cdots & \mathbf{Q}_{j,m} & \cdots & \mathbf{Q}_{j,M_{\text{modes}}} \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ \mathbf{Q}_{n_{\text{modes}},0} & \cdots & \mathbf{Q}_{n_{\text{modes}},m} & \cdots & \mathbf{Q}_{n_{\text{modes}},M_{\text{modes}}} \end{bmatrix}. \tag{44}$$

Here, $\mathbf{Q}_{j,m}$ is the matrix corresponding to the jth structural mode and the mth fluid mode, assembled over all the structural nodes and the fluid nodes at the interface Γ. The (A,C)th element of the matrix $\mathbf{Q}_{j,m}$ is given by

$$[\mathbf{Q}_{j,m}]_{AC}=\chi_{jm}\int_{\Gamma} N_C^f(x,0)N_A^s(x)\,dx, \tag{45}$$

where χ_{jm} is given in Eq. (16). Note that the lack of orthogonality between the fluid and structural modes leads to a $\mathbf{Q}$ that is full (as χ is not diagonal). The vectors $\mathbf{p}$, $\mathbf{d}$, and $\mathbf{f}$ are given by

$$\mathbf{p}=\begin{pmatrix} \mathbf{p}_0 \\ \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_{M_{\text{modes}}} \end{pmatrix}, \quad d=\begin{pmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \vdots \\ \mathbf{d}_{n_{\text{modes}}} \end{pmatrix}, \quad \mathbf{f}=\begin{pmatrix} \mathbf{f}_0 \\ \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_{M_{\text{modes}}} \end{pmatrix}. \tag{46}$$

For plane-wave input, as considered in this work, only $\mathbf{f}_0$ is nonzero. The Cth entry of the forcing vector $\mathbf{f}_0$ is given by

$$[\mathbf{f}_0]_C=\frac{-2ik\tilde{p}_{\text{in}}L_y}{\rho_f\omega^2}\int_{\Gamma_L} N_C^f(0,z)\,dz \tag{47}$$

if the corresponding node is along Γ_L. Otherwise, the forcing term is zero.

The matrix problem is solved for unknown structural displacements and fluid pressures from Eq. (39).

III. RESULTS AND DISCUSSION

The overall goals of this design are to maximize the transmission loss, and minimize the transmitted and reflected power over the frequency range of interest (100–2000 Hz). These noise-control characteristics are represented by the following three ratios: $\mathcal{TL}=-10\log_{10}(\mathcal{P}_t/\mathcal{P}_{\text{in}})$, $\mathcal{PR}_{\text{trans}}=\mathcal{P}_t/\mathcal{P}_{\text{in(net)}}$ and $\mathcal{PR}_{\text{refl}}=\mathcal{P}_{\text{refl}}/\mathcal{P}_{\text{in}}$, representing transmission loss, ratio of transmitted power to net input power, and ratio of reflected power to input power, respectively. Here, $\mathcal{P}_{\text{in}}$ is the power due to the plane-wave pressure input at $x=0$, $\mathcal{P}_t$ is the power transmitted across Γ_R into the domain Ω_R, $\mathcal{P}_{\text{in(net)}}$ is the net power transmitted across the boundary Γ_L into Ω, and $\mathcal{P}_{\text{refl}}$ is the power reflected back into the domain Ω_L. The three ratios are not independent as $\mathcal{PR}_{\text{trans}}=10^{-\mathcal{TL}/10}/(1-\mathcal{PR}_{\text{refl}})$. Therefore, only the transmission loss and the reflected power versus frequency are discussed in this paper.

The time averaged power transmitted through a surface Γ is computed as

$$\mathcal{P}=\frac{1}{2}\int_{\Gamma}\text{Re}(pv^*)\,d\Gamma, \tag{48}$$

TABLE I. Plate and fluid properties.

Property	Value	Units
Plate Young's modulus in transverse direction (E_1)	39	GPa
Plate Young's modulus in longitudinal direction (E_2)	8.6	GPa
Plate shear modulus (G)	3.8	GPa
Poisson's ratio of the plate (ν)	0.28	
Density of the plate (ρ)	2100	kg/m^3
Hysteritic damping factor of the plate (η)	0.05	
Density of water at room temperature (ρ_f)	1000	kg/m^3
Speed of sound in water at room temperature (c)	1500	m/s

where v^* is the complex-conjugate of velocity and p is the pressure. The power crossing the surfaces Γ, Γ_L, and Γ_R is computed by postprocessing the results of the finite element analysis.

A. Numerical experiments

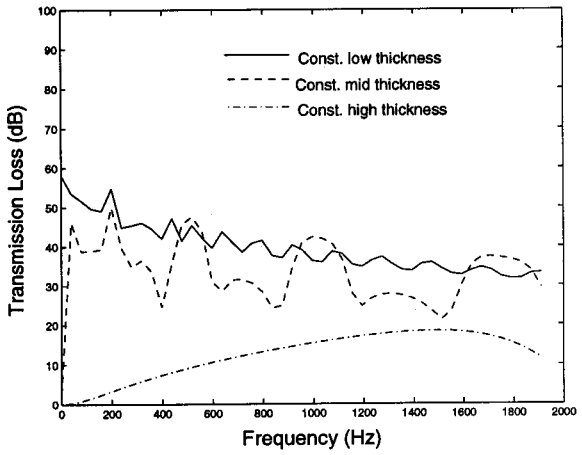
To analyze the noise-control performance of structural acoustic silencers, consider a structural layer comprised of E-glass epoxy plate of length 5 cm flush-mounted on a hydraulic pipe of rectangular cross section $3\times3\,\text{cm}^2$. The properties of the fluid and the plate (Humphreys and Rosen, 1993) are listed in Table I.

In the following subsections, the effect of spatial variations of the plate (width and thickness) and the degree of orthotropy on noise control performance are explored. Further, to compare with Huang's design concept (Huang, 1999; Huang *et al.*, 2000), the noise-control characteristics of plates and membranes are compared.

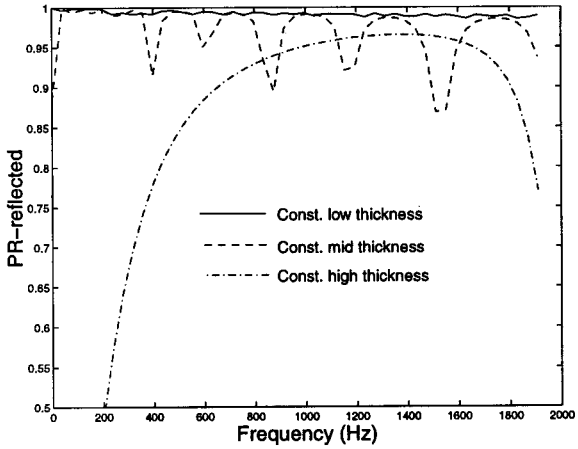
1. Constant thickness and width plate

The noise control characteristics of three plates of constant width ($b=3$ cm) with uniform thicknesses of $h=0.0012$ cm (low), 0.012 cm (mid), and 0.12 cm (high) are compared. The material properties are listed in Table I. The transmission loss and reflected power determined using power calculations are shown in Figs. 6(a) and (b).

It is observed that the plate of highest thickness provides the lowest transmission loss [Fig. 6(a)] while the plate of lowest thickness performs best with the highest transmission loss. The low thickness presents a high impedance discontinuity, thereby reflecting most of the energy back to the source. This is seen in Fig. 6(b), which shows the reflected power normalized to the input power. The increase in $\mathcal{TL}$ with frequency for constant high thickness plate is due to decrease in absolute value of the plate impedance with frequency. An opposite trend is observed for the case of constant low thickness plate which is due to an increase in the absolute value of the plate impedance with frequency for the particular set of parameters chosen in this numerical experiment. One drawback of the low thickness plate is that it can sustain only a low mean pressure. It is reasonable to expect that a plate of varying thickness and/or varying width would have better overall performance combining noise control and load bearing capacity.



(a) Transmission-loss



(b) Power-reflected normalized to power-input

FIG. 6. Plate of constant thickness and width. Plates of three different thicknesses, viz., 0.0012 cm (low), 0.012 cm (mid), and 0.12 cm (high) are considered. Plate width = 3 cm for the three cases. The orthotropic factor for the plates is 0.22.

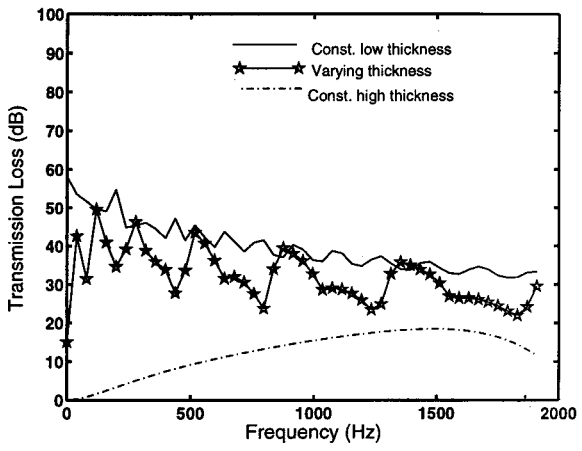
2. Varying impedance plate

Two types of varying impedance plates, one with varying thickness and the other with varying width, are analyzed (properties are listed in Table I). The plate stiffness decreases along the length in analogy with the cochlea. The dimensions of the varying thickness plate are computed from an approximate formula based on the assumption that the plate is strongly orthotropic (an array of parallel beams). Initial dimensioning is based on the fundamental *in vacuo* natural frequency, ω_n, of a beam with Young's modulus E_1, thickness h, length b, and density ρ under pinned boundary conditions, given by

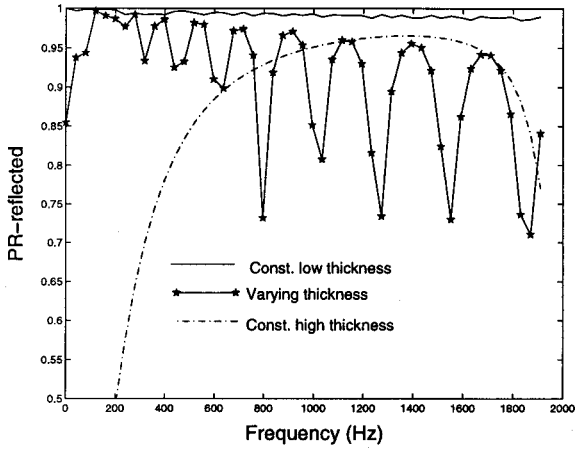
$$\omega_n = \frac{\pi^2}{b^2}\sqrt{\frac{E_1 h^3}{12\rho h}}. \tag{49}$$

In order to allow for deviations from the *in vacuo* frequency obtained above for nonstrongly orthotropic plates, the thickness and width for the varying impedance plates are obtained for a frequency range broader than the desired range. Equation (49) is used to determine the dimensions of a plate of constant width and varying thickness, $b = 3$ cm and $h(x) = 0.12e^{(-0.92x)}$ cm.

As might be expected, the performance of the varying thickness plate is midway between those of the constant impedance plates of low and high thicknesses. The varying thickness plate achieves higher transmission loss [see Fig. 7(a)] compared to the constant impedance plate of high thickness, and, over a few frequencies, out-performs the constant impedance plate of low thickness. Further, the varying thickness plate can take higher static pressure loading compared to the constant impedance plate of low thickness, thus making it a better candidate for noise control. Another advantage of the varying thickness plate over the constant impedance plate of low thickness is the reduced reflection as

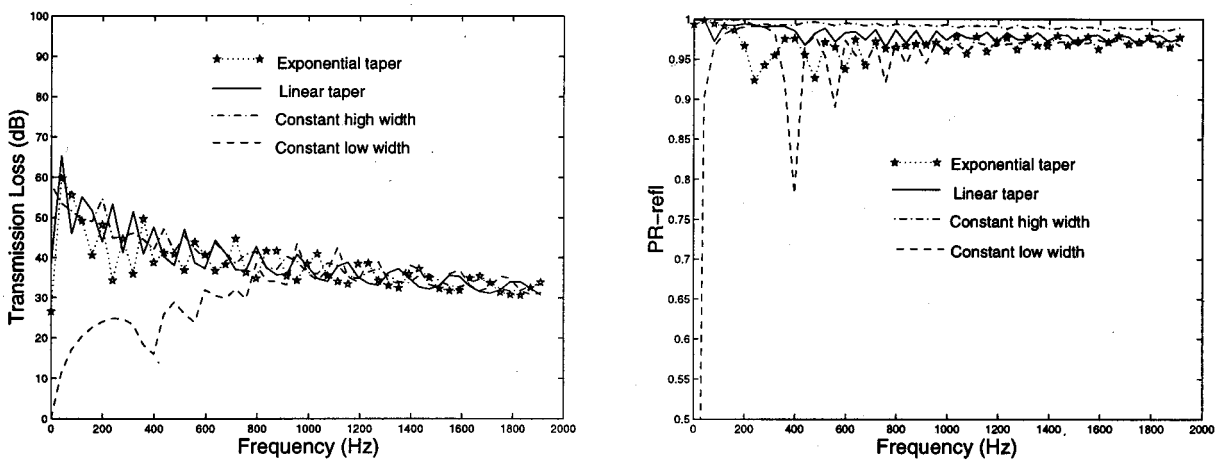


(a) Transmission loss



(b) Power-reflected normalized to power-input

FIG. 7. Plate of constant width and varying/constant thickness. Plate width = 3 cm (constant). Plate thickness (for varying thickness plate) varies as $h(x) = 0.12e^{(-0.92x)}$ cm. For constant thickness plates, the thicknesses are 0.0012 cm (low) and 0.12 cm (high). The orthotropic factor for the plates is 0.22.



(a) Transmission loss

(b) Power-reflected normalized to power-input

FIG. 8. Plate of constant thickness and varying/constant width. Plate thicknesses = 0.0012 cm. Plate width (for varying width plate) varies from 0.3 cm to 3 cm linearly/exponentially. Varying width plate is compared against constant width plates of width 0.3 cm (low) and 3 cm (high). The orthotropic factor for the plates is 0.22.

seen from Fig. 7(b). The analogy with cochlea is reflected by the fact that the peak of the velocity profile shifts leftward with frequency, the shift being sharp in the case of strongly orthotropic plate (figure not shown).

Next, a varying width plate is considered. In general, it is possible and advantageous to start from a zero width (lessening the impedance discontinuity). But for the purpose of comparison, the varying width plate is chosen to perform over the same frequency range as the varying thickness plate. The dimensions are obtained using Eq. (49). The plate width varies as $b = \epsilon L_y$, where ϵ varying from 0.1 to 1 is the fraction of the duct width occupied by the plate. The plate thickness is held constant with $h = 0.0012$ cm. For the given duct geometry, the width varies from 0.3 to 3 cm. Linear and exponential forms of width variation are considered. Results for two constant width and thickness plates are generated for comparison; the first, denoted as a low width plate, has a width of 0.3 cm corresponding to the narrow end of the varying width plate; the second, denoted as the high width plate, has a width of 3 cm corresponding to the widest end of the varying width plate. Note that the 3-cm-wide, 0.0012-cm-thick plate is same as the low thickness plate in Sec. III A 1.

From Fig. 8(a), the varying width plate yields higher transmission loss compared to the constant impedance plate of low width (with $b = 0.3$ cm and $h = 0.0012$ cm) and almost matches the transmission loss achieved by the constant impedance plate of high width (with $b = 3$ cm and $h = 0.0012$ cm).

Both linear and exponential variation in width provide transmission loss above 30 dB over the whole frequency range, while the transmission loss achieved by the plate with thickness variation falls below 30 dB for some frequencies. In Fig. 8(b), the reflected power for the varying width plate is shown to be smaller compared to the constant impedance plate of high width.

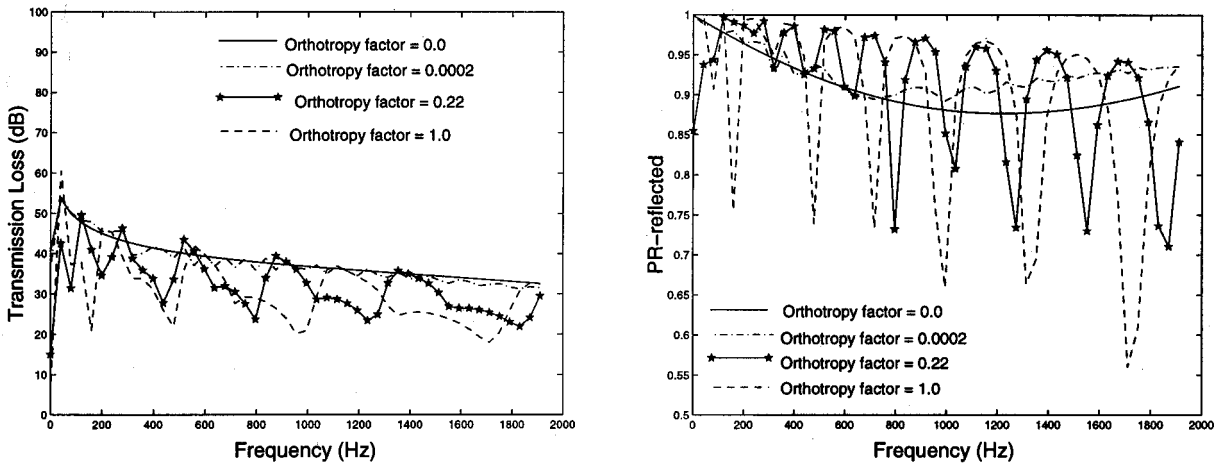
3. Effect of orthotropy

Motivated by the strongly orthotropic structure in the cochlea, it is of interest to find the effect of orthotropy on noise control. The material properties for the plate are modified from those listed in Table I so that E_1 is fixed at 39 GPa, ν is fixed at 0.28, E_2 is defined as ϱE_1, and G is defined as $E_2/2(1+\nu)$. Here, ϱ is the orthotropy factor that varies from 0 for a strongly orthotropic plate to 1 for an isotropic plate ($\varrho = 0.22$ for actual E-glass epoxy). In this numerical simulation, four different orthotropy factors, $\varrho = 0$, 0.0002, 0.22, and 1, are considered. The dimensions for thickness and width are obtained in the same fashion as in the case of varying thickness plate. The plate width is held constant with $b = 3$ cm and the thickness varies spatially as $h(x) = 0.12e^{(-0.92x)}$ cm. Results for the varying thickness case are presented here as the effect of orthotropy is seen more clearly than for the varying width case.

The transmission loss increases [Fig. 9(a)] with increasing relative stiffness in the transverse direction (i.e., $\mathcal{TL}$ increases as the plate becomes "more" orthotropic). The reflected power ratio shows a different trend. For an isotropic plate, the reflected power ratio reaches a value as low as 0.55 [Fig. 9(b)], whereas for a strongly orthotropic plate, the reflected power remains near 0.9 over most of the frequency range of interest.

4. Plate versus membrane

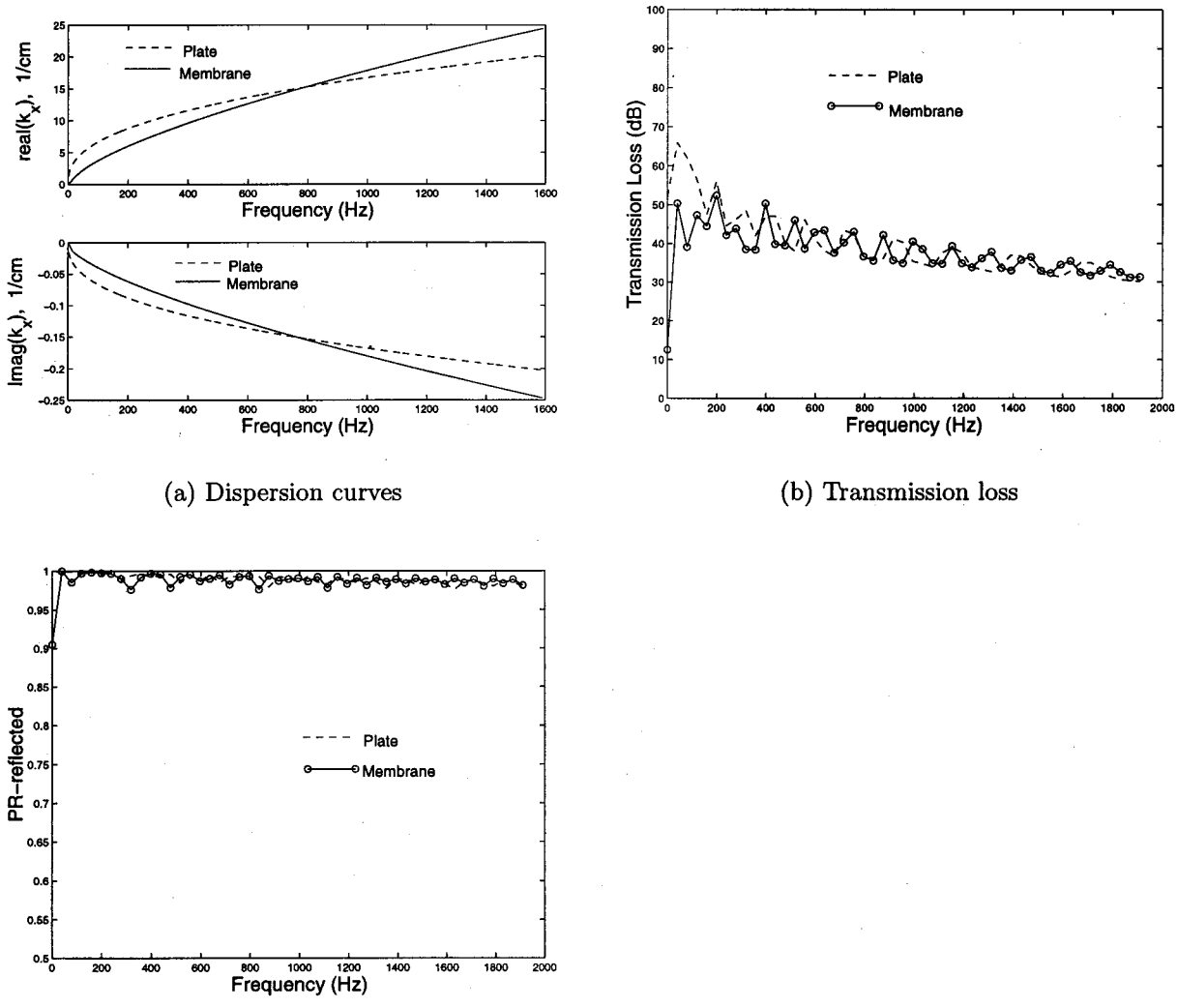
Since thin plates are used in this noise control device, it is of interest to study the effectiveness of a structural element that relies on tension rather than bending rigidity to act as the restoring force. In order to compare the noise control characteristics of a plate and a membrane, it is necessary to use a matching criterion. The hueristic criterion used in this paper is matching the wave number of the fluid-loaded plate and the membrane at a frequency midway through the frequency



(a) Transmission loss

(b) Power-reflected normalized to power-input

FIG. 9. Effect of orthotropy. Varying thickness plates of different orthotropic factors, viz., 0, 0.0002, 0.22, and 1 are considered. In each case, the plate width = 3 cm and plate thickness varies as $h(x) = 0.12e^{(-0.92x)}$ cm.



(a) Dispersion curves

(b) Transmission loss

(c) Power-reflected normalized to power-input

FIG. 10. Effect of dispersion mechanism. Isotropic plate and isotropic membrane are compared for noise control performance. The isotropic membrane has the same wave number as the isotropic plate at 800 Hz (properties of such a membrane, given the isotropic plate, are derived using dispersion analysis). Plate/membrane width = 3 cm and thickness = 0.0012 cm.

TABLE II. Membrane properties.

Property	Value	Units
Applied Tension per unit membrane width (T_0)	14.088	N/m
Mass per unit area (ρh)	0.0252	kg/m^2
Membrane wave speed (c_m)	23.7	m/s
Damping factor (η)	0.03	

range of interest so that performance above and below the matching frequency can be examined. For simplicity, an isotropic constant impedance plate is compared with an isotropic membrane of constant impedance by matching the wave numbers at 800 Hz. The properties of the plate are similar to those listed in Table I, with the only change that the Young's modulus E_2 is set equal to E_1, i.e., $E_2=E_1=39$ GPa. The constant impedance plate has the dimensions $b=3$ cm and $h=0.0012$ cm. The isotropic membrane (with uniform applied tension) of constant impedance, as determined by matching the wave number at 800 Hz with that of the plate [Fig. 10(a)], has the properties listed in the Table II.

The noise control characteristics of the two structures individually in contact with water are compared over the whole frequency range of interest [Figs. 10(b) and (c)].

Since both the plate and the membrane represent a strong impedance mismatch, the transmission loss for both the cases is very high. However, the plate has better overall performance for frequencies below the matching frequency, where it achieves higher transmission loss compared to the membrane.

B. Interpretation of the results using wave number analysis

The free response of a fluid-loaded plate can be analyzed using the Wentzel–Kramer–Brillouin (WKB) approximation to understand the results of the FEA. The WKB method is an asymptotic method for approximate solution to differential equations involving a slowly varying parameter. The use of the WKB method in this paper is necessitated by the variable impedance plate. The form of the plate displacement (and fluid pressure) used for the WKB method is $w(x,y,z)=A(x)e^{i\phi(x)}\psi(y,z)$, where $A(x)$ is varying amplitude, $\phi(x)=\int k_x(x)\,dx$ is the accumulated phase ($k_x=k_{xr}+ik_{xi}$ is the plate wave number), and $\psi(y,z)$ is the modal solution in y and z directions, x being the direction along which the plate material properties vary. For details on WKB methodology, see Steele and Taber (1979) and Parthasarathi (2000). Even for a constant impedance plate, the WKB formulation gives physical insight into the wave number aspect of the problem. The WKB formulation leads to an *eikonal* equation for the wave number and a transport equation for the plate displacement. As a simplified example, consider a one-dimensional approximation for the fluid and a strongly orthotropic plate, which has the following closed form eikonal equation (see, e.g., deBoer, 1984; Parthasarathi, 2000)

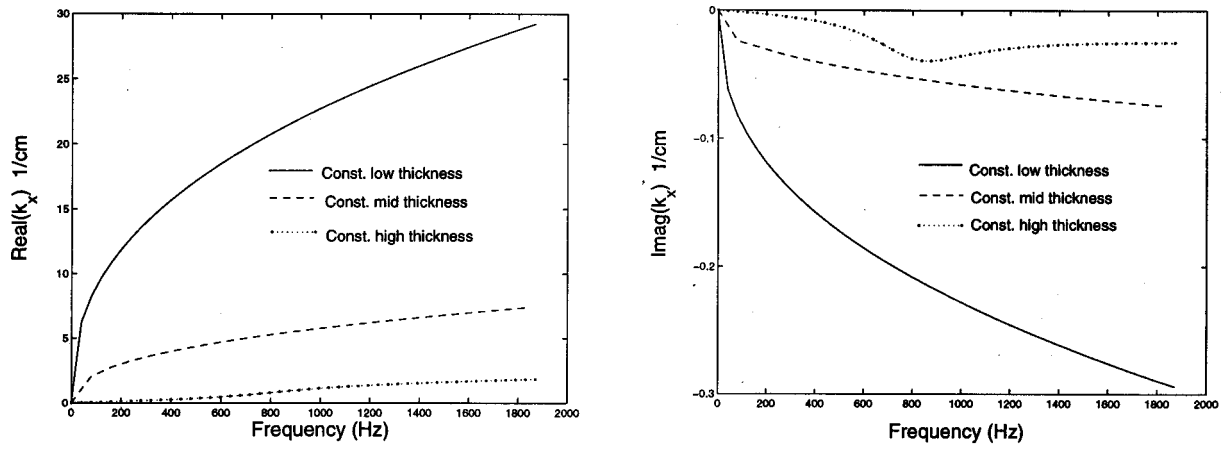
$$1+\frac{\pi^2 L_z Z(x)}{16i\omega\rho_f\epsilon}k_x^2=0, \tag{50}$$

where

$$Z(x)=\frac{1}{i\omega}\frac{E_1h^3}{12(1-\nu^2)}\left(\frac{\pi}{b}\right)^4-i\omega\rho h \tag{51}$$

is the plate impedance. From this relation, it is noted that the wave number is inversely proportional to the plate impedance. The plots of wave number versus frequency generated using eikonal equation in 3-d for constant impedance plates with $\varrho=0.22$ discussed in Sec. III A 1 are shown in Figs. 11(a) and (b). Note that no closed form solution of 3-d eikonal equation for k_x is possible.

High value of k_{xi} implies rapid decay of the plate velocity. Since plate velocity determines the scattered pressure (Huang, 1999), the oscillations in scattered pressure and hence the oscillations in total pressure are reduced when k_{xi} is high. When k_{xi} is low, this implies that the structural waves propagate along the entire length of the plate, allowing for constructive and destructive interference which, in turn, leads to fluctuations of $\mathcal{TL}$. Figure 11(b) indicates that k_{xi} is high for constant low impedance plate, which explains



(a) Real part of the plate wavenumber

(b) Imaginary part of the plate wavenumber

FIG. 11. WKB results for variation of wave number with plate impedance. Plates of three different thicknesses, viz., 0.0012 cm (low), 0.012 cm (mid), and 0.12 cm (high), are considered. Plate width=3 cm for the three cases. The orthotropic factor for the plates is 0.22.

the reduced oscillations in $\mathcal{TL}$ in Fig. 6(a) compared to constant mid impedance plate. Along the same lines of argument, if only the plate orthotropy is varied keeping the transverse direction Young's modulus constant, impedance increases as orthotropy decreases (as ϱ increases), leading to increased amplitude of $\mathcal{TL}$ oscillations [Fig. 9(a)].

IV. CONCLUSIONS AND FUTURE WORK

From the current analysis, it is seen that a structural acoustic silencer indeed achieves high transmission loss over a broad band of frequencies. In this analysis, the acoustic wavelengths are much longer than the dimensions of the structural acoustic silencer. The length of structural acoustic silencers that achieve 30–60 dB transmission loss are less than 1% of the wavelength at 200 Hz.

The effectiveness of plate orthotropy for noise control is quite evident from Fig. 9(a). Relative stiffening of the plate in the transverse direction makes it a better silencer. Varying the impedance provides for excellent noise control while simulatenously increasing the tolerable static pressure load, as compared to constant thin plates. More effective handling of static pressure is an important area of future work. Multiduct designs that separate the variable impedance layer from the static pressure loads and are still effective for noise control remain a challenge (Fig. 3). A key issue is the rejection of the plane wave mode that would propagate through the duct without exciting the structural layer. For the plate designs used in this paper, the dominant contribution to high transmission loss arises from reflection due to the impedance discontinuity rather than absorption in the plate. However, the results for varying impedance suggest the possibility of reducing reflections and increasing the transmission loss by using a plate with simultaneous thickness and width variations along the length. This impedance tailoring reduces the impedance jump felt by an incoming wave when it reaches the structural acoustic silencer. The design concept of a varying impedance plate allows for frequency tailored filtering of acoustic systems and broadband noise control.

ACKNOWLEDGMENTS

This research is funded by the Office of Naval Research and the National Science Foundation. The assistance of Dr. Anand A. Parthasarathi in modeling and computer code is acknowledged.

Dallos, P., Popper, A., and Fay, R. (**1996**). *The Cochlea* (Springer-Verlag, New York).

deBoer, E. (**1984**). ''Auditory physics. Physical principles in hearing theory. II,'' Phys. Rep. **105**, 141–226.

Dodson, J. M. (**2001**). ''Efficient Finite Element Methods/Reduced-Order Modeling for Structural Acoustics with Applications to Transduction,'' dissertations, University of Michigan.

Dodson, J. M., Dowling, D. R., and Grosh, K. (**1998**). ''Experimental investigation of quarter wavelength silencers in large-scale hydraulic systems,'' Noise Control Eng. J. **46**, 15–22.

Dodson, J. M., and Grosh, K. (**under revision, 2002**). ''A 2.5d finite element method for fluid-loaded, variable-width, orthotropic plates: theory and experiment,'' J. Acoust. Soc. Am.

Edge, K. A., and Johnston, D. N. (**1990a**). ''The 'secondary source' method for the measurement of pump pressure ripple characteristics part 1: description of the method,'' Proc. Inst. Mech. Eng., Part A: J. Power Energy **204**, 32–40.

Edge, K. A., and Johnston, D. N. (**1990b**). ''The 'secondary source' method for the measurement of pump pressure ripple characteristics part 2: experimental results,'' Proc. Inst. Mech. Eng. Part A: J. Power Energy **204**, 41–46.

Golub, G., and Loan, C. V. (**1989**). *Matrix Computations* (Johns Hopkins, Baltimore, MD).

Huang, L. (**1999**). ''A theoretical study of duct noise control by flexible panels,'' J. Acoust. Soc. Am. **106**, 1801–1809.

Huang, L. (**2001**). ''A theoretical study of passive control of duct noise using panels of varying compliance,'' J. Acoust. Soc. Am. **109**, 2805–2814.

Huang, L., Choy, Y. S., So, R. M. C., and Chong, T. L. (**2000**). ''Experimental study of sound propagation in a flexible duct,'' J. Acoust. Soc. Am. **108**, 624–631.

Hughes, T. J. R. (**1987**). *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis* (Prentice–Hall, Englewood Cliffs, NJ).

Humphreys, E. A., and Rosen, B. W. (**1993**). ''Properties analysis of laminates,'' in *Engineered Materials Handbook, Vol. 1: Composites* (ASM International).

Martin, H. R. (**1984**). ''Noise control in hydraulic systems,'' *SAE Technical Paper Series, 841062, International Off-Highway and Powerplant Congress and Exposition*.

Munjal, M. L. (**1987**). *Acoustics of Ducts and Mufflers* (Wiley, New York).

Munjal, M. L., and Thawani, P. T. (**1996**). ''Acoustic performance of hoses—a parametric study,'' Noise Control Eng. J. **44**, 274–280.

Parthasarathi, A., and Grosh, K. (**2000**). ''Three-dimensional numerical modeling for global cochlear dynamics,'' J. Acoust. Soc. Am. **107**, 474–485.

Parthasarathi, A. A. (**2000**). ''Numerical modeling and electro-acoustic stimulus response analysis for cochlear mechanics,'' dissertation, University of Michigan.

Pierce, A. D. (**1989**). *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY).

Ramamoorthy, S., and Grosh, K. (**2000**). ''Broadband structural acoustic silencers,'' J. Acoust. Soc. Am. **108**, 2464.

Steele, C., and Taber, L. (**1979**). ''Comparison of wkb calculations and experimental results for three-dimensional cochlear models,'' J. Acoust. Soc. Am. **65**, 1007–1018.

Strunk, R. D. (**1991**). ''Silencer for hydraulic piston pump pressure pulsations,'' *SAE Technical Paper Series, 911759, International Off-Highway and Powerplant Congress and Exposition*.

Yu, J., Johnson, F., Iwami, F., Verrecchia, N., and Kojima, E. (**1999**). ''Experimental evaluation for fluidborne noise attenuation in tuning cables and hoses of automotive power steering hydraulic systems,'' *SAE Technical Paper Series, 1999-01-1777*.

Hearing protector attenuation: Models of attenuation distributions

William J. Murphy[a] and John R. Franks
Hearing Loss Prevention Section, Engineering and Physical Hazards Branch, Division of Applied Research and Technology, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, MS C-27, Cincinnati, Ohio 45226-1998

Edward F. Krieg
Monitoring Research and Statistics Activity, Division of Applied Research and Technology, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, MS R-3, Cincinnati, Ohio 45226-1998

(Received 2 April 2001; revised 20 September 2001; accepted 21 December 2001)

Current hearing protector rating standards estimate the protection performance for a given frequency as the mean attenuation minus a multiple of the standard deviation. Distributions of real-ear attenuation at threshold data are fit with maximum likelihood estimation procedures using both normal and mixed-normal models. Attenuations from six hearing protectors, Bilsom UF-1 earmuff, Bilsom Quietzone, E•A•R® Classic®, E•A•R® EXPRESS® Pod Plugs®, Howard Leight MAX, and Wilson EP100 earplugs, measured with a subject-fit protocol are reported. The mixed-normal (bimodal) model provides a better fit to the empirical data than the unimodal model for most frequencies and protectors. Primarily, the bimodal model fits the shape of the distributions caused by data from poorly-fit protectors. This paper presents an alternative method for estimating the protection performance either with the more accurate bimodal model or directly from the empirical cumulative distributions of the attenuation data. [DOI: 10.1121/1.1461835]

PACS numbers: 43.50.Hg, 43.66.Vt, 43.15.+s [MRS]

I. INTRODUCTION

Real-ear attenuation at threshold (REAT) determines hearing protector attenuation as the difference between occluded and unoccluded thresholds for third-octave band noise stimuli at specified center frequencies from 63 to 8000 Hz depending upon the standard used. Three hearing protector fitting protocols have been defined in ANSI standards: experimenter-fit in ANSI S3.19-1974 (ANSI, 1974), experimenter-supervised-fit, and subject-fit in ANSI S12.6-1997 (ANSI, 1997).

ANSI S3.19-1974 provided for an experimenter fit when optimum protector performance was required. The subject was to fit the hearing protector, however the experimenter inspected the fit to assure a good fit and acoustic seal. The experimenter had the option of reinserting earplugs or readjusting other protectors to achieve a ''best'' fit. This method afforded the greatest control over the REAT testing. In fact, Franks *et al.* (1996) demonstrated that EF data exhibited the smallest lab-to-lab, subject-to-subject, and trial-to-trial variance while achieving the greatest mean attenuations. However, EF data have significantly overestimated real-world data as indicated by the derating of the noise reduction rating (NRR) by OSHA inspectors (Murphy and Franks, 1998, 2001; OSHA, 1983, 1999).

Similarly, the experimenter-supervised-fit (ESF) protocol described in Method A of ANSI S12.6-1997 also overestimates the real-world attenuation (Royster *et al.*, 1996; Berger *et al.*, 1998). The ESF protocol yields the greatest lab-to-lab variation, and has a tendency to include more data from subjects with poorly-fit protectors (Murphy and Franks, 1998). The European community uses a protocol similar to the ESF protocol described in ISO 4869-1 (1990) for estimating REAT for protectors. Calculation of the ISO Single Number Rating (SNR) and HML are accomplished with the REAT data collected under an ESF protocol. ANSI S3.19-1974, Method A of ANSI S12.6-1997, and the ISO ESF protocols prescribe the use of a fitting noise which aids the subject in achieving the best fit of the protector.

The subject-fit (SF) protocol (ANSI S12.6-1997, Method B) gives the subject control of fitting the device. Subjects tested under the SF protocol are required to be naïve with respect to hearing protector use and training. When tested, the subject is given a copy of the manufacturer's instructions and instructed that the experimenter is not allowed to assist in the fitting process. The SF data typically have the lowest mean attenuations of the three test protocols and the greatest subject-to-subject and trial-to-trial variances. The SF lab-to-lab variances were less than the ESF lab-to-lab variances and were comparable to the variances for the EF method (Murphy and Franks, 1998). The agreement between SF data and real-world data demonstrated that naïve subject performance was comparable to workers who were supposedly trained[1] in using earplugs (Berger *et al.*, 1998). The primary shortcoming with the subject-fit protocol is the increased prevalence of poorly-fit protectors. Murphy and Franks (1998) examined the distribution of attenuations from four protectors tested in interlaboratory studies using EF, ESF, and SF protocols and found that the premolded plugs, Bilsom Quietzone and Wilson EP100, exhibited bimodal dis-

[a]Electronic mail: wjm4@cdc.gov

tributions for several test frequencies. The bimodal attenuation distributions often had a peak near 0 dB and another at about 20 dB which typified poorly-fit and well-fit protectors.

The NRR, SNR, high middle low (HML) and the noise reduction rating subject-fit (NRR_{SF}) are among several rating methods currently in use in the United States and Europe. The 1995 NIOSH Hearing Protector Compendium lists the rating method in use at that time. For details about the methods, one should consult the compendium and the appropriate standard (NIOSH, 1995; EPA, 1978; ISO, 1994; Berger and Royster, 1996; Franks *et al.*, 2000). In essence, the ratings are subtracted from the noise levels measured for a worker's station and yield an estimate of the noise levels under the protector (ANSI S12.19, 1996). Hearing protectors sold in the United States are required to be labeled under the EPA rule, 40 CFR 211, subpart B. The EPA rule requires that testing be done according to the EF protocol in the ANSI S3.19-1974 standard even though ANSI S12.6-1997 supercedes S3.19. European standards require that protectors be tested according to ISO 4869 part 1 and labeled according to ISO 4869 parts 2 and 3 which specify the SNR and HML rating methods.

One common element to each of these protected noise level calculations is the use of the mean attenuation minus some multiple of the standard deviation. Hence, the attenuation values are assumed to be normally distributed when describing the percentage of a protected population. In Murphy and Franks (1998), the low-frequency attenuation distributions for the Bilsom Quietzone and Wilson EP100 earplugs were non-normal and bimodal while the distributions for the E•A•R® Classic® and Bilsom UF-1 were unimodal and normal. This trend was most pronounced for the subject-fit data.

For this paper, the data measured in several laboratories for the four hearing protectors examined in Murphy and Franks (1998) and the data from two additional devices, the E•A•R® EXPRESS® Pod Plugs® and Howard Leight MAX earplugs, were analyzed. The assumption of normality was critically examined to determine its appropriateness as a descriptor of REAT data. The goal was to develop a more representative estimate of protection afforded by hearing protectors which can be used in current rating standards with minimal modification of the calculations of the rating values.

II. THEORY FOR BIMODAL DATA

The developments of the various rating methods have been predicated upon the assumption that data are well behaved. Attenuations for a hearing protector measured at several different frequencies are described by means and standard deviations. Consistent attenuation data are assumed to be normally distributed or at the least unimodal such that the protector has one way to fit the ear canal and that the continuum of attenuations measured at several frequencies are best characterized by a mean and standard deviation. Subject-fit test protocols provide more opportunity for the test subjects to improperly fit the protector to the ear. When the protector is improperly fit, the integrity of the protector's seal with the head can be compromised resulting in low attenuations due to leakage. Because SF protocols prohibit experimenter intervention to correct improper fits, the REAT measurements will exhibit more variability than ESF or EF protocols.

A. The Gaussian distribution

When sampling a population, the samples are assumed to be drawn randomly from a normally-distributed population,

$$g(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}, \tag{1}$$

where x is a continuous variable, μ is the mean, and σ is the standard deviation. The Gaussian distribution is important to the hearing protector rating standards because multiples of the standard deviation are used to estimate the protection performance. Specifically, the protection performance for NRR, SNR, HML, and NRR_{SF} standards is

$$\text{Protection Performance}\,(\alpha)=\int_{\mu-\alpha\sigma}^{\infty}g(x)\,dx, \tag{2}$$

where $\alpha=2$ for the NRR, $\alpha=1$ for the NRR_{SF}, and α is variable but is normally set to 1 for the SNR and HML methods.

B. Bimodal attenuation distribution

A mixture of two Gaussian distributions can be used to create a bimodal distribution. Since no *a priori* knowledge of the bimodality is assumed, the data are apportioned between the two distributions. The combined distribution becomes

$$f(x)=\left(\frac{\phi}{\sqrt{2\pi}\sigma_1}e^{-(x-\mu_1)^2/2\sigma_1^2}+\frac{1-\phi}{\sqrt{2\pi}\sigma_2}e^{-(x-\mu_2)^2/2\sigma_2^2}\right), \tag{3}$$

where μ_1 and μ_2 are the means, σ_1 and σ_2 are the standard deviations, and ϕ is a proportionality constant with a range of 0 to 1.

Integration of $f(x)$ yields the bimodal cumulative distribution

$$F(x)=\phi\int_{\mu-\alpha\sigma}^{\infty}g(x,\mu_1,\sigma_1)dx + (1-\phi)\int_{\mu-\alpha\sigma}^{\infty}g(x,\mu_2,\sigma_2)\,dx, \tag{4}$$

where $g(x,\mu,\sigma)$ is the Gaussian function in Eq. (1). Figure 1 illustrates the mixed Gaussian distribution and its associated cumulative distribution function. The bimodal cumulative distribution has a characteristic shoulder between the two peaks. If the two distributions overlap considerably; the mixture of two normal distributions can look like a single, skewed distribution, or even resemble a single normal distribution.

III. METHODS

The data used in these analyses have been previously reported in several articles: Royster *et al.* (1996), Franks

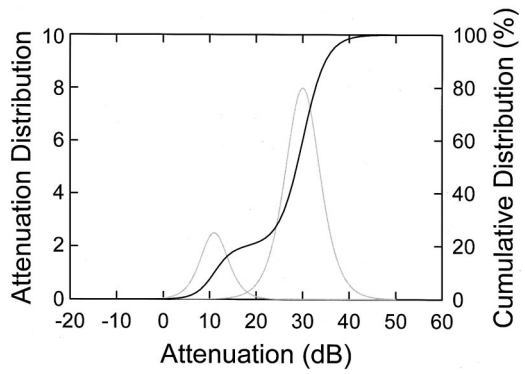


FIG. 1. The bimodal cumulative distribution derived from two normal Gaussian distributions. Different proportionality constants, means, and standard deviations affect the shoulder in this plot.

et al. (1996), Berger *et al.* (1998), and Franks *et al.* (2000). Data were collected at E•A•RCal laboratory, Wright Patterson Air Force Base, US Army Auditory Research Labs, Virginia Tech University, and at the NIOSH Taft Laboratories. Each of the studies collected and analyzed REAT data using the EF, ESF, or SF protocol(s). This paper analyzes the subject-fit data collected by these laboratories for the Bilsom UF-1 earmuff (UF-1), Bilsom Quietzone (Quietzone), E•A•R® Classic® (Classic), E•A•R® EXPRESS® Pod Plugs® (Express), Howard Leight MAX (MAX), and Wilson EP100 (EP100) earplugs.

The E•A•RCal, WPAFB, USAARL, and NIOSH laboratories tested the UF-1, Classic, V-51R, and EP100 hearing protectors with 24 subjects each and four REAT repetitions per subject in the four-lab interlaboratory study (Royster *et al.*, 1996) for a total of 96 subjects. In the two-lab study, Virginia Tech and NIOSH tested Classic and V-51R hearing protectors with 26 and 25 subjects, respectively, with three repetitions per subject in the two-lab study (Murphy and Franks, 1998) for a total of 51 subjects. The MAX earplug was tested with 24 subjects with three repetitions per subject and 24 subjects with two repetitions per subject. The Express earplug was tested with 20 subjects with two repetitions per subject (Franks *et al.*, 2000). Both the MAX and Express devices were tested at NIOSH. In order to make the data compliant with the ANSI S12.6 Method B protocol, the first two repetitions were used in the analysis. Within a frequency band, the REAT estimates were averaged over the first two trials to create a data set which did not have repeated measures for a subject. In Royster *et al.* (1996), the REAT data exhibited a lab-to-lab effect. While the effect is real, it was not subtracted from the laboratory averages because of the possibility that it resulted from different relative distributions of the well- and poorly-fit data.

Subjects were recruited from the communities surrounding the labs and were compensated for the participation. Subjects were prescreened to determine their lack of experience with hearing protector usage, training, and testing. They were trained to produce sound-field audiograms with a 6-dB range or better for each test frequency for three consecutive audiograms. Diffuse sound-field tests were conducted in reverberant chambers that satisfied the ANSI S12.6-1997 requirements. Specifically, the sound field should vary no more than ± 5 dB at a radius of 15 cm about the center of the subject's head and no more than 3 dB for the left–right axis. Pairs of unoccluded and occluded thresholds were measured for each subject. Between trials, subjects removed the hearing protection and, if using earplugs, were given a new pair of plugs to use on the next occluded trial.

IV. RESULTS

The REAT data were fit with both the single and mixed Gaussian models with the general optimization function **ms** in S-Plus (Venables and Ripley, 1997). The means (μ_1, μ_2), standard deviations (σ_1, σ_2), and proportionality constants (ϕ) are reported in Table I. The fits were then tested with a likelihood ratio test to determine whether the bimodal model fit was significantly better than the unimodal fit. A p value less than 0.05 was the criterion for statistical significance. Table II lists the p values for each of the comparisons of the fits.

The data in Table I exhibit some interesting trends. The proportionality constant tends to be low for those frequencies and protectors where the unimodal model gives a reasonable fit. The exception to this trend is the UF-1 earmuff. In cases where the two distributions overlap significantly, the mean for unimodal model is between the means of the bimodal fit, and the unimodal standard deviation is comparable to those fit for the bimodal model.

For the insert earplugs, $0.1 < \phi < 0.9$ tends to indicate bimodally-distributed data. For the Express at 125 Hz, the high proportionality constant, 0.913, resulted from the fitting optimization. The formulation of the theory is symmetric with respect to ϕ. A shoulder at about 0.1 can be seen in the 125-Hz Express panel in Fig. 2. This shoulder is also evident in the other panels for the Express. In all of these cases, the fitting algorithm captured the feature.

In Fig. 2, the cumulative probability distributions of the data, unimodal, and bimodal Gaussian models (y axis) are plotted against the cumulative probability of the unimodal Gaussian model (x axis). In this plot, the straight line is the normal probability plot that the data would overlay if it were normally distributed. The bimodal fit is plotted with a gray line. In several cases, notably the Quietzone and EP100, deviations about the straight line indicate non-normal behavior. Most of the deviations occur in the lower left portion of each panel which represent the minimum REAT for that given device and frequency.

The maximum-likelihood test of the unimodal and bimodal fit demonstrated that the bimodal model yielded a better fit in 27 of 42 cases. The bimodal model did not provide poor fits to the normally-distributed data, the fits were just not significantly better than the unimodal fits. Clearly, the Quietzone and EP100 data in Fig. 2 demonstrated greater propensity for bimodality at the low frequencies. At 2000 Hz and above, the low attenuation shoulders in the distributions were diminished. The bimodal fit to the Quietzone data was not significantly better at 4000 Hz than the unimodal fit. The Express cumulative distributions exhibited a variety of shapes. At 125 Hz, the distribution was normally distributed. At the other frequencies, the distributions appeared to have a shoulder at the low attenuations. The bimodal fits for 125

TABLE I. Fitting results for unimodal and bimodal models of attenuation.

Protector Model		Bilsom UF-1		E•A•R® Classic®		Wilson EP100		Howard Leight MAX		E•A•R® EXPRESS® Pod Plugs®		Bilsom Quietzone	
		Uni	Bi	Uni	Bi	Uni	Bi	Uni	Bi	Uni	Bi	Uni	Bi
125	μ_1	7.4	4.2	20.8	9.0	14.4	1.4	21.8	1.7	14.2	12.7	12.5	2.1
Hz	σ_1	3.6	2.0	6.8	5.4	11.0	2.4	8.2	0.9	6.9	5.3	9.5	2.4
	μ_2		8.5		21.9		20.9		23.1		29.5		17.5
	σ_2		3.3		5.8		7.1		6.7		2.6		7.3
	ϕ		0.265		0.084		0.336		0.060		0.913		0.324
250	μ_1	14.2	11.3	21.6	10.0	14.3	1.9	22.4	1.8	15.2	3.6	11.9	1.7
Hz	σ_1	29	3.6	6.3	4.3	10.7	2.5	8.8	1.1	6.6	2.5	9.1	2.3
	μ_2		14.7		22.1		20.9		23.7		18.0		16.6
	σ_2		2.3		5.9		6.7		7.4		3.4		7.0
	ϕ		0.163		0.039		0.349		0.059		0.198		0.313
500	μ_1	20.8	16.5	23.5	19.4	15.0	0.8	23.5	0.4	16.5	14.3	12.4	0.7
Hz	σ_1	2.9	1.8	7.2	0.8	11.7	1.7	9.7	0.1	7.7	9.4	9.1	1.4
	μ_2		21.9		23.8		20.0		24.5		19.0		15.2
	σ_2		2.1		7.3		9.4		8.6		3.5		7.9
	ϕ		0.193		0.063		0.259		0.041		0.542		0.194
1000	μ_1	29.4	20.4	24.7	12.3	16.8	2.4	23.7	2.8	19.0	4.7	14.7	2.5
Hz	σ_1	4.0	1.3	6.2	0.4	10.4	3.2	7.5	1.6	8.0	3.8	9.1	2.8
	μ_2		30.2		25.0		21.8		24.6		22.6		18.8
	σ_2		3.1		6.0		6.8		6.2		3.6		6.9
	ϕ		0.081		0.019		0.257		0.040		0.198		0.232
2000	μ_1	31.6	31.5	31.0	19.1	23.9	13.3	30.2	11.7	27.4	17.0	22.3	7.5
Hz	σ_1	4.1	3.2	4.6	1.0	9.9	7.6	6.0	3.6	8.8	9.9	9.4	4.4
	μ_2		32.0		31.2		29.3		31.0		31.6		26.0
	σ_2		6.0		4.3		5.8		4.7		2.4		6.1
	ϕ		0.763		0.024		0.334		0.042		0.291		0.202
4000	μ_1	35.8	34.2	39.0	33.5	29.9	23.2	38.6	31.2	32.1	17.4	23.1	8.7
Hz	σ_1	4.0	4.0	5.6	6.5	10.5	8.6	7.1	7.2	7.7	5.7	7.5	2.1
	μ_2		37.6		40.9		39.0		42.1		35.0		23.9
	σ_2		3.2		3.8		3.9		3.3		3.6		6.9
	ϕ		0.524		0.249		0.576		0.322		0.167		0.054
8000	μ_1	35.2	25.1	38.4	33.6	26.8	10.9	39.7	32.3	31.2	9.7	18.9	10.6
Hz	σ_1	4.5	1.7	7.4	7.1	13.5	5.6	8.5	9.5	9.0	4.5	10.7	5.3
	μ_2		36.0		42.3		34.9		43.9		33.5		24.2
	σ_2		3.6		4.9		8.1		3.8		5.5		9.9
	ϕ		0.070		0.452		0.336		0.357		0.100		0.392

and 500 Hz were not significantly better than the unimodal fits. The fewer number of data points for the Express yield a rougher data plot and gave the appearance of a greater amount of noise in the data than for the EP100 and Quietzone devices.

The Classic data were more normally distributed. At 4000 Hz, the Classic data had a slight shoulder at the lower attenuations. Only for the 4000 and 8000 Hz distributions were the bimodal fits significantly better than the unimodal fits. The cumulative distributions for the MAX earplug were more noisy than the Classic due to fewer data points. The MAX cumulative distributions had a small shoulder at 1000 Hz and below. At 4000 and 8000 Hz, the distributions deviated from the unimodal fit in the middle sections. The bimo-

TABLE II. The results of maximum likelihood ratio tests. The *p* values which are less than 0.05 are in bold type and indicate that the bimodal model provides a better fit than the unimodal model.

	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Bilsom Quietzone	$\mathbf{6.46\times10^{-9}}$	$\mathbf{1.60\times10^{-8}}$	$\mathbf{2.21\times10^{-7}}$	$\mathbf{6.69\times10^{-5}}$	$\mathbf{2.10\times10^{-5}}$	3.11×10^{-1}	$\mathbf{6.25\times10^{-3}}$
Bilsom UF-1	4.43×10^{-1}	1.31×10^{-1}	7.47×10^{-2}	$\mathbf{3.22\times10^{-3}}$	3.89×10^{-1}	8.03×10^{-1}	$\mathbf{2.15\times10^{-2}}$
E•A•R® Classic®	2.34×10^{-1}	8.89×10^{-1}	3.33×10^{-1}	4.74×10^{-1}	1.75×10^{-1}	$\mathbf{1.92\times10^{-4}}$	$\mathbf{4.90\times10^{-2}}$
E•A•R® EXPRESS®	4.07×10^{-1}	$\mathbf{2.50\times10^{-2}}$	5.32×10^{-1}	$\mathbf{5.83\times10^{-3}}$	$\mathbf{4.12\times10^{-4}}$	$\mathbf{1.02\times10^{-2}}$	$\mathbf{5.42\times10^{-2}}$
Howard Leight MAX	$\mathbf{1.21\times10^{-2}}$	5.15×10^{-2}	$\mathbf{7.18\times10^{-3}}$	7.61×10^{-2}	$\mathbf{1.84\times10^{-2}}$	$\mathbf{1.65\times10^{-3}}$	$\mathbf{9.67\times10^{-4}}$
Wilson EP100	$\mathbf{1.72\times10^{-9}}$	$\mathbf{3.54\times10^{-9}}$	$\mathbf{5.88\times10^{-8}}$	$\mathbf{3.10\times10^{-5}}$	$\mathbf{1.40\times10^{-2}}$	$\mathbf{2.47\times10^{-4}}$	$\mathbf{1.39\times10^{-4}}$

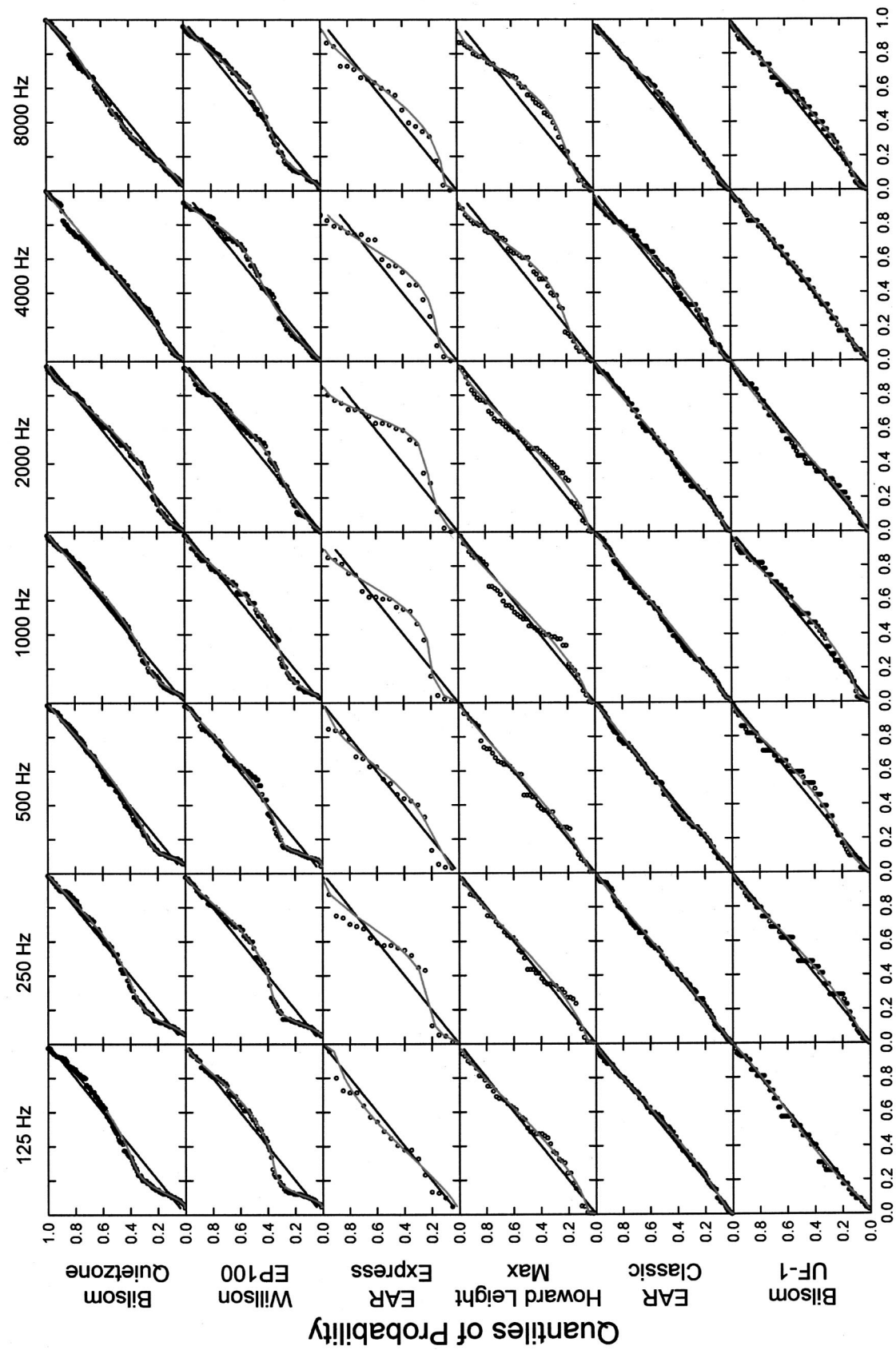


FIG. 2. Normal probability plots for six hearing protectors measured at 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. The abscissa coordinates are the unimodal Gaussian quantiles. The ordinate coordinates are the probability of the cumulative distribution. The empirical data are shown as circles. By definition, the unimodal Gaussian fits are the straight black lines. The bimodal fits are the gray lines.

TABLE III. Comparison of the 16th percentile derived from the empirical data, and the bimodal and unimodal models. The values in the table are in dB and represent attenuations associated with each protector at the beginning of the rows.

	Model	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Bilsom	Empirical	1.6	1.5	1.5	3.5	10.9	15.7	8.6
Quietzone	Bimodal	1.9	1.6	1.3	3.5	10.7	15.5	8.2
	Unimodal	3.1	2.9	3.3	5.6	12.9	15.6	8.2
Bilsom	Empirical	4.0	12.0	17.6	25.6	28.5	32.0	31.1
UF-1	Bimodal	3.7	11.6	17.7	25.9	27.9	31.8	31.3
	Unimodal	3.8	11.3	17.9	25.4	27.6	31.8	30.8
E•A•R®	Empirical	14.3	15.7	17.0	19.0	27.0	34.3	31.3
Classic®	Bimodal	14.4	15.4	16.8	18.6	26.6	34.0	30.7
	Unimodal	14.1	15.3	16.4	18.6	26.4	33.5	31.0
E•A•R®	Empirical	6.4	7.0	8.0	11.0	22.3	27.1	26.8
EXPRESS® Pod Plugs®	Bimodal	7.8	5.8	9.2	8.0	18.2	25.8	25.3
	Unimodal	7.3	8.6	8.8	11.1	18.7	24.4	22.2
Howard Leight	Empirical	15.2	16.9	16.2	17.8	25.8	31.3	32.3
MAX	Bimodal	14.8	14.6	14.6	17.4	25.6	31.2	31.1
	Unimodal	13.6	13.7	13.9	16.3	24.2	31.6	31.3
Wilson	Empirical	1.5	1.6	1.1	2.9	11.6	18.0	10.3
EP100	Bimodal	1.2	1.6	1.0	3.3	12.8	18.1	10.5
	Unimodal	3.5	3.7	3.4	6.4	14.1	19.4	13.4

dal fits were significantly better at 125, 500, 2000, 4000, and 8000 Hz.

The UF-1 data exhibited smooth responses at all frequencies which were close to the normally-distributed ideal. The bimodal model was judged more likely than the unimodal only at 8000 Hz.

For each frequency, a linear interpolation of the data points bracketing the 16th percentile (0.1573) was used to estimate equivalent $\mu-1\sigma$ of the empirical cumulative distribution. The 16th percentile of the unimodal model was calculated directly as $\mu-1\sigma$. For the bimodal model, the cumulative distribution was numerically solved for the attenuation at the 16th percentile. The results for the six devices are compared in Table III. For the Quietzone, large discrepancies between the unimodal and the empirical estimates of the 16th percentile were evident below 4000 Hz. The bimodal model was within 0.3 dB of the empirical estimate at all frequencies. The bimodal estimate was closer than the unimodal at all frequencies except 4000 Hz where the unimodal was 0.1 dB less than the empirical.

Similarly for the EP100, the bimodal estimates were closer to the empirical data than the unimodal at all frequencies. The unimodal estimates overestimated the 16th percentile by as much as 3.5 dB at 1000 Hz. At 2000 Hz, the bimodal differed by 1.2 dB while the unimodal differed by 2.4 dB.

The best estimates for the Express were mixed between the bimodal and the unimodal models. At 250, 4000, and 8000 Hz, the bimodal model gave a better estimate of the empirical data. At 1000 Hz, the difference between the bimodal model and the data was 3.0 dB. The unimodal model gave a better estimate of the data at 125 and 2000 Hz even though the data were apparently distributed bimodally. If additional data were collected for the Express, the cumulative distributions would be expected to exhibit less noise and would potentially be better fit by both the bimodal and unimodal models.

The bimodal model 16th percentile estimates for the Classic fits exhibited good agreement with the data. At 125 Hz, the bimodal estimate differed by 0.1 dB and was within 0.4 dB at 500 and 1000 Hz. At other frequencies, the bimodal model was within 0.1 dB. The unimodal model differed by 0.8 dB at 2000 Hz and 0.7 dB at 4000 Hz. At other frequencies, the unimodal fits were within 0.5 dB of the empirical data.

The unimodal fits to the MAX data exhibited more discrepancy than the corresponding fits to the Classic data. At 125 Hz, the unimodal fit was off by 0.7 dB while the bimodal fits were off by 1.6 dB at all other frequencies. The bimodal model provided a better estimate of the 16th percentile.

Lastly, the bimodal model provided a better fit than the unimodal model at most of the frequencies for the UF-1 data. At 4000 Hz, the two estimates were the same. At 125 and 1000 Hz, the unimodal estimate was 0.1 dB better than the bimodal estimate. The bimodal estimate was never more than 0.6 dB different from the empirical estimate while the unimodal estimate was 0.9 dB less than the empirical estimate at 2000 Hz.

V. DISCUSSION

The unimodal and bimodal models presented in this paper have been examined for how well they describe subject-fit REAT data. The unimodal model is limited in its ability to describe the results, and yet the unimodal Gaussian distribution forms the basis of data analysis for the current rating standards and rules. In particular, the SNR rating method (ISO 4869−2) calculates the assumed protected value as

$\mu - \alpha\sigma$, where α is variable. The NRR uses $\alpha = 2$ and the NRR_{SF} uses $\alpha = 1$. While the standards never explicitly state the assumption of normality, it is implicit in the treatment of the data. Therefore, the validity of the assumption should be tested before applying the Gaussian method. Murphy and Franks (1998) explicitly tested normality and demonstrated the V51R and EP100 data were not normally distributed.

The use of subject-fit protocols almost certainly guarantees that future hearing protector REAT data will include results from poorly-fit protectors. Consequently, standards which assume well-behaved data require revision to include testing of the assumption of normality and need to accurately model the data. The analysis of these data demonstrates the bimodal model yields better fits in most cases and does a better job of predicting the assumed protected values at the 16th percentile.

One difference between ANSI and ISO subject-fit protocols should be noted. The ANSI S12.6-1997 Method B specifically requires that the test subjects be naive with respect to hearing protector use and training. The ISO standard has no such requirement and, consequently, the experience with hearing protector use is uncontrolled. In particular, the poor performance of some earplugs can be remedied by training subjects in proper insertion technique and how to use fitting noise to identify a poorly sealed protector. However, the ISO 4869-1 values tend to overpredict the real world attenuation of hearing protectors whereas the Method B data fall within the upper quartile of real world measurements (Berger *et al.*, 1998).

The analysis presented here is not specifically dependent upon the use of a Gaussian distribution. Earlier formulations (Murphy and Franks, 1999a, 1999b) utilized the logistic distribution to model the bimodal distributions. The analytic solution for the cumulative distribution yielded a simpler analysis. Other statistical distributions may prove to be more appropriate for fitting skewed data distributions. The unimodal Gaussian distribution has been critically examined and found inadequate when the attenuation data are contaminated with poorly-fit protectors.

VI. CONCLUSIONS

This paper has sought to develop a better approach to model the attenuation of hearing protectors. With the advent of new standards which utilize the subject-fit protocol, testers have relinquished control over how test subjects will fit the hearing protector. Similarly, the manufacturer has no control over how the workers wear their hearing protectors. Because of this fundamental change in the control of data integrity, the assumption of normality is not guaranteed. Therefore, the bimodal model addresses the problem of non-normal distributions of REAT data and more accurately models the variety of data. The standards-setting bodies, ANSI and ISO, have the responsibility to develop scientifically defensible standards. Future ANSI and ISO standards for rating the noise reduction of hearing protectors should be based on accurate models of the attenuation distribution. Determination of the protection performance directly from the cumulative distribution or from the bimodal fit to the data would remedy this shortcoming in the current HPD rating standards. At the very least, ISO and ANSI standards could be based on empirical quantiles which do not assume any particular attenuation distribution.

[1]OSHA's construction standard, 29 CFR 1926.101(b), titled: ''Hearing Protection,'' requires ''ear protective devices inserted in the ear shall be fitted or determined individually by competent persons.'' Workers are typically given little training in the proper use of insert earplugs as can be seen by visual inspection of workers's earplugs at numerous occupational settings.

ANSI (**1974**). S3.19-1974, ''American National Standard Method for the Measurement of Real-Ear Protection of Hearing Protectors and Physical Attenuation of Earmuffs,'' (American National Standards Institute, New York).

ANSI (**1997**). S12.6-1997, ''American National Standard Method for Measuring Real-Ear Attenuation of Hearing Protectors,'' (American National Standards Institute, New York).

ANSI (**1996**). S12.19-1996, ''American National Standard Measurement of Occupational Noise Exposure,'' (American National Standards Institute, New York).

ANSI (**1995**). S12.42-1995, ''American National Standard Method for Measuring Microphone in Real-Ear Technique,'' (American National Standards Institute, New York).

Berger, E. H., Franks, J. R., and Lindgren, F. (**1996**). ''International review of field studies of hearing protector attenuation,'' in *Scientific Basis of Noise-Induced Hearing Loss*, edited by A. Axelsson, H. Borchgrevink, R. P. Hamernik, L. Hellstrom, D. Henderson, and R. J. Salvi (Thieme, New York), pp. 361–377.

Berger, E. H. and Royster, L. H. (**1996**). ''In search of meaningful measures of hearing protector effectiveness,'' Spectrum Suppl. 1, **13**, 29.

Berger, E. H., Franks, J. R., Behar, A., Casali, J. G., Dixon-Ernst, C., Kieper, R. W., Merry, C. J., Mozo, B. T., Nixon, C. W., Ohlin, D., Royster, J. D., and Royster, L. H. (**1998**). ''Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices. Part III. The validity of using subject-fit data,'' J. Acoust. Soc. Am. **103**, 665–672.

Code of Federal Regulations, Labor, Chapter 29, Part 1926, Paragraph 101(b) (**1999**). ''Hearing Protection,'' revised July 1, 1999, the Office of the Federal Register, National Archives and Records Administration, U.S. Government Printing Office, Washington, DC 20204-9328.

Franks, J. R., Murphy, W. J., and Simon, S. D. (**1996**). ''Repeatability and reproducibility in hearing protector testing,'' J. Acoust. Soc. Am. **99**, 2464.

Franks, J. R., Murphy, W. J., Johnson, J. L., and Harris, D. A. (**2000**). ''Four earplugs in search of a rating system,'' Ear Hear. **21**, 218–226.

ISO 4869-1 (**1990**). ''Acoustics—Hearing Protectors—Part 1: Subjective method for the measurement of sound attenuation,'' (International Organization for Standardization, Geneva).

ISO 4869-2 (**1994**). ''Acoustics—Hearing Protectors—Part 2: Estimation of Effective A-Weighted Sound Pressure Levels When Hearing Protectors are Worn,'' (International Organization for Standardization, Geneva).

ISO/TR 4869-3 (**1989**). ''Acoustics—Hearing Protectors—Part 3: Simplified Method for the Measurement of Insertion Loss of Ear-Muff Type Protectors for Quality Inspection Purposes,'' (International Organization for Standardization, Geneva).

Murphy, W. J., and Franks, J. R. (**1998**). ''Analysis of Repeatability and Reproducibility of Hearing Protector Rear-Ear-Attenuation-at-Threshold Measured with Three Fitting Methods,'' National Hearing Conservation Association, 19–21 February, Albuquerque, NM.

Murphy, W. J., and Franks, J. R. (**1999a**). ''A rating system in search of protectors,'' J. Acoust. Soc. Am. **105**, 1131(A).

Murphy, W. J., and Franks, J. R. (**1999b**). ''Progress on a rating system for hearing protector attenuation,'' J. Acoust. Soc. Am. **106**, 2262(A).

Murphy, W. J., and Franks, J. R. (**2001**). ''A reevaluation of the Noise Reduction Rating,'' Commissioned Officers Association of the US Public Health Service, Washington, DC, 29 May 2001.

NIOSH (**1995**). *The NIOSH Compendium of Hearing Protection Devices*, U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Publication No. 95-105.

OSHA (**1983**). *Guidelines for Noise Enforcement: Appendix A*, U.S. Department of Labor, Occupational Safety and Health Administration, OSHA Instruction CPL 2-2.335A.

OSHA (**1999**). *OSHA Technical Manual, Section III, Health Hazards, Chapter 5, Noise Measurement*, U.S. Department of Labor, Occupational Safety and Health Administration, OSHA Instruction TED 1-0.15A.

Royster, L. H. (**1995**). ''Recommendations for Labeling Hearing Protectors.'' Sound Vib. **29**, 16–19.

Royster, J. D., Berger, E. H., Merry, C. J., Nixon, C. W., Franks, J. R., Behar, A., Casali, J. G., Dixon-Ernst, C., Kieper, R. W., Mozo, B. T., Ohlin, D., and Royster, L. H. (**1996**). ''Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices. Part I. Research of Working Group 11, Accredited Standards Committee S12, Noise,'' J. Acoust. Soc. Am. **99**, 1506–1526.

Venables, W. N., and Ripley, B. D. (**1997**). *Modern Applied Statistics with S-Plus*, 2nd ed. (Springer-Verlag, New York), Sec. 9.7.

U.S. Environmental Protection Agency (**1978**). CFR Title 40, sub chapter G, 211, subpart B-Hearing Protective Devices.

Nondestructive imaging of shallow buried objects using acoustic computed tomography

Waheed A. Younis
Department of Electrical and Computer Engineering, University of Western Ontario, London, Ontario N6A 5B9, Canada

Stergios Stergiopoulos[a)]
Defence and Civil Institute of Environmental Medicine, Toronto, Ontario M3M 3B9, Canada and Department of Electrical and Computer Engineering, University of Western Ontario, London, Ontario N6A 5B9, Canada

David Havelock
Institute for Microstructural Sciences, National Research Council, Ottawa, Ontario K1A 0R6, Canada

Julius Grodski
Defence and Civil Institute of Environmental Medicine, Toronto, Ontario M3M 3B9, Canada

(Received 11 June 2001; accepted for publication 19 February 2002)

The nondestructive three-dimensional acoustic tomography concept of the present investigation combines computerized tomography image reconstruction algorithms using acoustic diffracting waves together with depth information to produce a three-dimensional (3D) image of an underground section. The approach illuminates the underground area of interest with acoustic plane waves of frequencies 200–3000 Hz. For each transmitted pulse, the reflected-refracted signals are received by a line array of acoustic sensors located at a diametrically opposite point from the acoustic source line array. For a stratified underground medium and for a given depth, which is represented by a time delay in the received signal, a horizontal tomographic 2D image is reconstructed from the received projections. Integration of the depth dependent sequence of cross-sectional reconstructed images provides a complete three-dimensional overview of the inspected terrain. The method has been tested with an experimental system that consists of a line array of four-acoustic sources, providing plane waves, and a receiving line array of 32-acoustic sensors. The results indicate both the potential and the challenges facing the new methodology. Suggestions are made for improved performance, including an adaptive noise cancellation scheme and a numerical interpolation technique.
[DOI: 10.1121/1.1470507]

PACS numbers: 43.58.Ta, 43.38.Hz, 43.40.Le [SLE]

I. INTRODUCTION

Acoustical techniques have been used to do subsurface image reconstruction for imaging and classifying buried objects for the past two decades. Most of these techniques use either well-to-well tomography[1] or surface-to-well tomography.[2–4] However, their methodology is not appropriate for detecting buried land mines or hazardous waste materials, since they require ground disturbances that could be hazardous for demining and destructive for archeological applications.

A recent study[5] investigates two-dimensional (2D) imaging of shallow buried objects by using ultrasound B-scan tomography.[6] The system concept includes a linear array of receivers and transmitters operating in the frequency range of (1–5) kHz. In this case, three-dimensional (3D) imaging can be accomplished by volume rendering of consecutive 2D images.[7] If the linear sensor array is replaced by a planar array, then 3D beamforming techniques[6] could provide the volume visualization of the underground area without combining multiple 2D images.

Advancements in remote sensing have led to the development of nondestructive subsurface detection methods such as infrared (IR) imaging,[8,10] ground penetrating radar (GPR),[9,10] seismic refraction, electromagnetic sensing[10] and electrical conductivity. All these techniques use perturbations in seismic or electromagnetic waves to detect, locate, and identify the buried objects. Currently, GPR and IR imaging are widely used in demining applications. The performance of GPR for detecting landmines degrades considerably with increased moisture in the ground and, like the electromagnetic sensors, it fails to detect nonmetallic objects. The IR imaging methods can only detect recently buried (1 month period) objects. The poor performance characteristics of the existing nondestructive imaging techniques can be improved by using data and image fusion techniques in a multisensor system that has shown to have the potential for reliable mine detection and classification.[10]

Another challenge of international significance is the disposal of buried hazardous waste, which requires reliable estimation of the location and nature of the buried

[a)]Author to whom correspondence should be addressed. Electronic mail: stergios@dciem.dnd.ca

substances.[11] For archeologists, subsurface imaging techniques are useful for assessing the historical significance of a site before beginning costly digging.[12]

The present paper considers an alternative approach to nondestructive 3D imaging by using the concept of computed tomography (CT), which is successfully used for noninvasive medical imaging diagnostic applications.[13] Briefly, medical CT imaging uses x rays to obtain cross-sectional images, or ''slices,'' of the human body and image reconstruction algorithms that are based on the Radon theorem.[14,15] In this investigation we propose subsurface image reconstruction of horizontal cross sections at different depths. Once such images of sufficient quality are available, a next step could include implementation of pattern recognition algorithms to identify and classify the mines. An advantage of our approach is the ability to detect mines that are buried beneath one another, as this is a common deployment procedure to deceive the conventional mine detection systems. The method is demonstrated with an experimental setup, which illustrates both the challenges and opportunities for imaging of shallow buried objects using acoustic computed tomography.

II. PRINCIPLES OF OPERATION OF ACOUSTIC CT

The basic principles of a CT data acquisition process and the relevant image reconstruction algorithms are discussed in detail in Refs. 3, 13–19. The data acquisition process consist of taking projection measurements of the object of interest defined by the image $f(x,y)$. The projection data form a 2D array called sinogram $p(r_n,\theta_i)$ $(n=1,...,N)$ $(i=1,...,M)$ where θ_i is the projection angle and r_n is the distance of nth detector from the center of the field of view.[13] N and M are the total number of detectors in the array and the total number of projections, respectively. This set of projections is used by CT image reconstruction algorithms to reconstruct the image $f(x,y)$. If the energy used to illuminate the object is nondiffracting (e.g., x rays), the image reconstruction algorithms are, *iterative algebraic techniques, Fourier slice theorem techniques* and *filtered back projection*.[13–15,17] Whereas for diffracting energy (e.g., sound waves), these algorithms are *filtered back propagation and Fourier diffraction projection techniques*.[3,16,18,19]

In this investigation we propose to implement the CT concept using acoustic energy to reconstruct subsurface images of horizontal cross sections of the entire volume of a buried object at different depths. We assume that the received signal carries information from different ground penetration depths that are represented as different time delays, indicated by the temporal index j below,

$$p(r_n,\theta_i,t_j) \quad (n=1,...,N), \quad (i=1,...,M), \quad (j=1,...,K). \tag{1}$$

As a result, for a given time delay t_0, image reconstruction of the projection data set $p(r_n,\theta_i,t_0)$ will provide horizontal tomography images of the buried object of interest corresponding to a specific depth in the ground. Then, repetition of the image reconstruction process for all the time delays $(j=1,...,K)$ would define a 3D volume consisting of horizontal underground tomography images. However, the correspondence between time delay t_0 and specific depth in the ground requires stratification of the underground medium. For a nonstratified medium, such as the experimental set up of this study discussed next, the above correspondence between time delay and depth is not valid due to the multipath effects of the acoustic waves in the homogenous underground.

Figure 1 depicts the experimental setup of the proposed acoustic tomography concept for nondestructive imaging of shallow buried objects. The acoustic sensors (microphone array) and source array, shown in Fig. 1, do not have a contact with the surface of the ground. Thus, the whole system could be mounted on a platform that may be deployed by an armored vehicle or is remotely controlled. For each transmitted acoustic pulse, the receiving array collects the signal, which is reflected–refracted from the internal underground buried objects and the ground inhomogenouties.

Therefore, the proposed approach addresses the inverse problem, which is the underground 2D and 3D imaging. Parameters related to the propagation problem (forward problem)

the physics of acoustic/elastic wave propagating in a three-dimensional structure,

what waves are expected to generate in the geometry of the experimental setup,

what are their expected arrivals, their travel time and their amplitude?

are not required by the proposed approach to address the imaging problem. Although, solutions to the forward problem can provide the essential parameters for a better image reconstruction approach,[20] they impose a practical implementation problem for a number of nondestructive imaging underground applications. In particular, to identify the 3D propagation characteristics, it would require significant amount of information in terms of boundary conditions, density estimates, and stratification characteristics. This kind of information, although essential for an optimum approach, makes the acoustic tomography concept impractical for demining or archeological applications because it requires ground disturbances to collect the essential information. Our goal and objectives of this experimental study is the development of an acoustic CT technique for nondestructive imaging of underground features that is practically realizable. However, the theoretical and experimental aspects relevant with the propagation problem (forward problem) are discussed in Sec. IV B of this paper, since they are essential to verify our experimental results.

III. PRACTICAL ISSUES AND PROPAGATION CHARACTERISTICS

During the course of the present investigation we were confronted with several challenges. First and foremost was the issue that the propagation characteristics of the acoustic signals in this setting were not known. More specifically, what is received by the receiving microphone array, depicted in Fig. 1, is a superposition of direct waves (direct arrivals through air), surface waves, subsurface waves, reflected–refracted energy from buried objects, reflected energy from

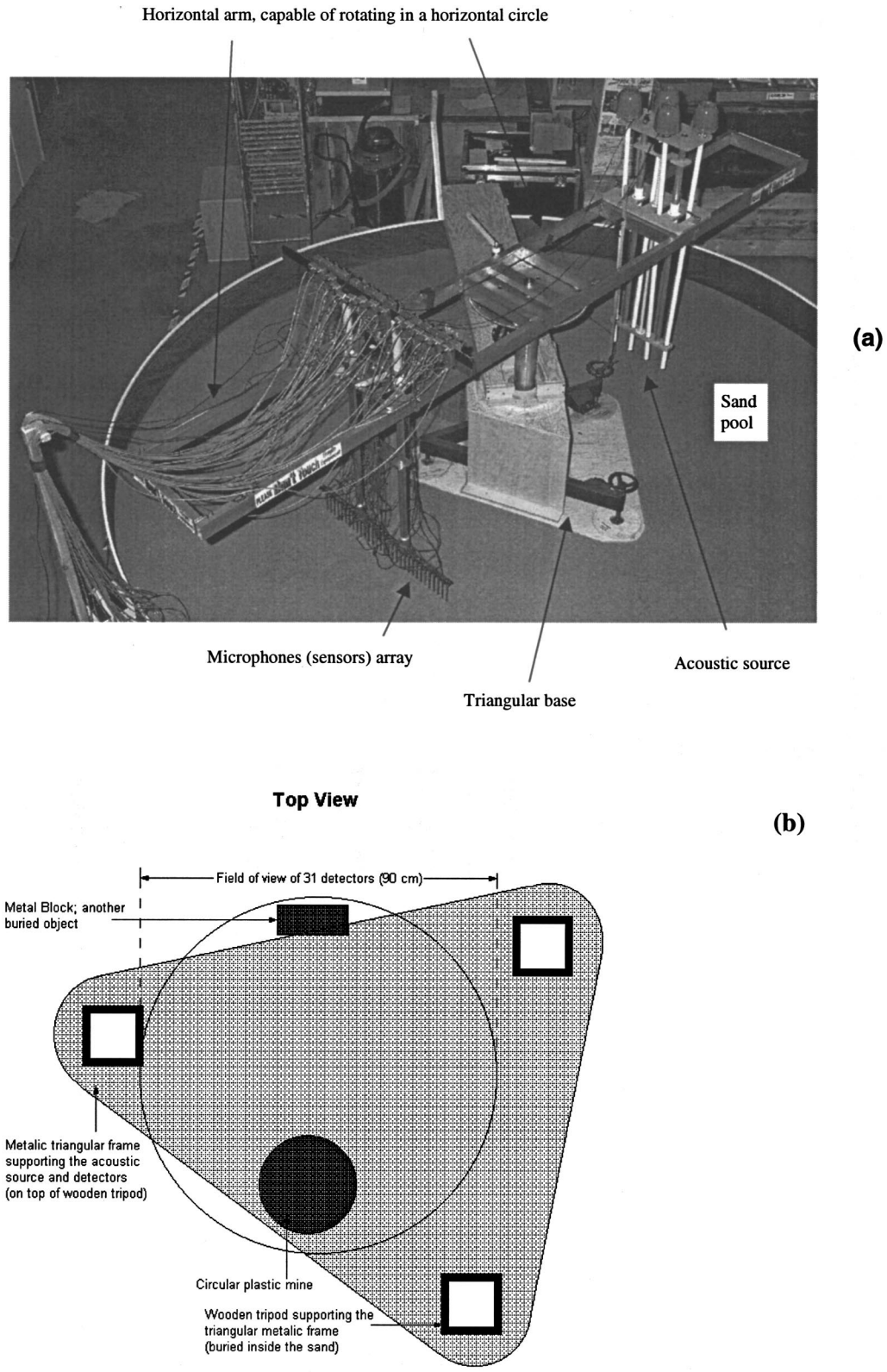


FIG. 1. (a: Top) Experimental setup of the acoustic CT concept including the receiving 32-microphone array and the four-speaker source array. (b: Bottom) Top schematic view of the experimental setup. The dark features indicate buried objects. The triangular base was on the surface of the sand. The circle indicates the field of view of the 90 cm line array receiver during the CT data acquisition process.

various inhomogeneous ground layers and noise. Each received acoustic wave front has a different time delay depending upon the propagation path it followed and the speed of sound in various sections of that path. Because of the poor air-to-ground-to-air coupling interface, another problem is to receive sufficient acoustic energy out of the soil by the receiving array[5] that is essential for reliable image reconstruction. The soil is a highly attenuating medium and it is diffi-

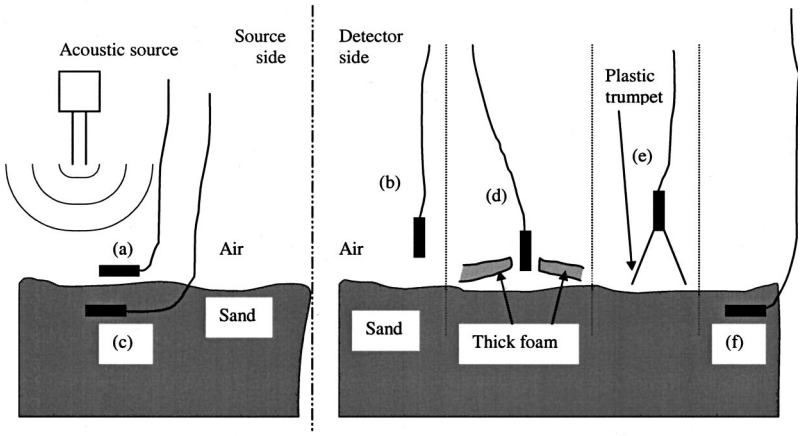


FIG. 2. Experimental setup of microphones to examine propagation effects for signals shown in Fig. 3.

cult for acoustic waves in the frequency range of 1–5 kHz to penetrate deep in the ground.[21] Although high frequency acoustic waves (2–10 MHz) produce good image resolution in ultrasound imaging, they are not suitable for underground imaging applications due to their severe attenuation in the ground. Thus a lower frequency range of 0.2–3.0 kHz was selected due to its characteristics to penetrate in the ground, though it produces poor image resolution. Furthermore, we constructed a frequency-modulated FM pulse, linearly varying from 200 Hz to 2000 Hz, cascaded with a Parzen window. In order to generate a plane wave acoustic signal to illuminate the underground area of interest, a linear array of four synchronized acoustic point sources was used, as shown in Fig. 1.

The poor signal-to-noise-ratio (SNR) in the received signal, as mentioned above, was mainly due to the poor air-to-ground-to-air coupling interface that induced a severe attenuation in the signal of interest. To minimize the impact of the very poor SNR in the image reconstruction process, we introduced an adaptive noise (interference) cancellation (AIC)[6,22,23] processing scheme. Its impact is demonstrated with real data in Sec. IV D.

IV. EXPERIMENTS

A. Experimental setup

The proposed acoustic CT tomography imaging system was implemented as shown in Fig. 1(a). A minefield was simulated by filling a commercial swimming pool of 4 m diameter and 0.7 m depth with dry construction grade sand. A wooden tripod structure with three hollow wooden boxes buried in the sand [Fig. 1(b)] was built to support the data acquisition system above the sand. The data acquisition system consisted of a 31-microphone receiving line array (Panasonic broad band capsule model WM063P) with 3 cm spacing and a four-speaker line array, mounted at the two sides of a horizontal metal bar (3.5 meter long). Two objects (simulating landmines) were buried in the sand [Fig. 1(b)] at different depths in the range of 10 cm to 50 cm: (1) an inert antivehicle plastic landmine 30 cm in diameter and (2) a metal flat box $5\times20\times30\ \mathrm{cm}^3$. A set of 32 preamplifiers (amplification=40 dB) was placed in the close vicinity of the microphones [Fig. 1(a)]. A 4 kHz, 32-channel antialiasing filter was used to condition the received signal for the 32-channel 12-bit A/DC (analog-to-digital converter with 8 kHz sampling frequency). The digitized time series from the 32-channel were multiplexed and stored in a SCSI storage device. The CT data acquisition process included 360 set of measurements at 1-degree interval around a horizontal circle defined by the rotated metal bar with the mounted receiving and source arrays [Fig. 1(a)]. For each set of measurements, the four-speaker line array generated a plane wave pulse sent in the ground; and the data acquisition system was triggered simultaneously to acquire the reflected acoustic signals by the 31-microphone receiving array.

To align all the received time series from the 360 measurements, a reference microphone (32nd sensor) was placed near the four-speaker array. This procedure of temporal alignment, using the received signal from the reference microphone, will be discussed later in Sec. IV C.

The transmitted acoustic signal was a FM pulse generated by a PC through a D/AC (digital-to-analog converter) and the audio amplifiers of the four-speaker array (Fig. 1). The FM pulse's characteristics were another major design issue, since they affect the image resolution. A very short pulse results in better image resolution and better differentiation of signal arrivals from different directions. However, the constraints imposed by the frequency range of 200–2000 Hz did not allow for the design of a very short FM pulse. A trial and error process provided a practical choice of a 0.1 ms FM pulse with 100 kHz sampling for the D/AC unit.

B. Propagation characteristics of the acoustic signals in the underground medium

A set of experiments was conducted to determine the propagation speed, time of arrival, amplitude and SNR of the propagating acoustic signals in the air and in the ground, as these were important parameters to determine the sections of the received microphone time series that included the signals of interest.

The sections (a)–(f) of Fig. 2 and (a)–(f) of Fig. 3 show the various experimental arrangements in terms of microphone positions and the corresponding received signals, respectively. More specifically, in Fig. 2, the microphone (a) is the reference microphone placed near the acoustic source; 2 meters away is the microphone (b), which is one of the microphones of the 31 sensor receiving array. The microphones

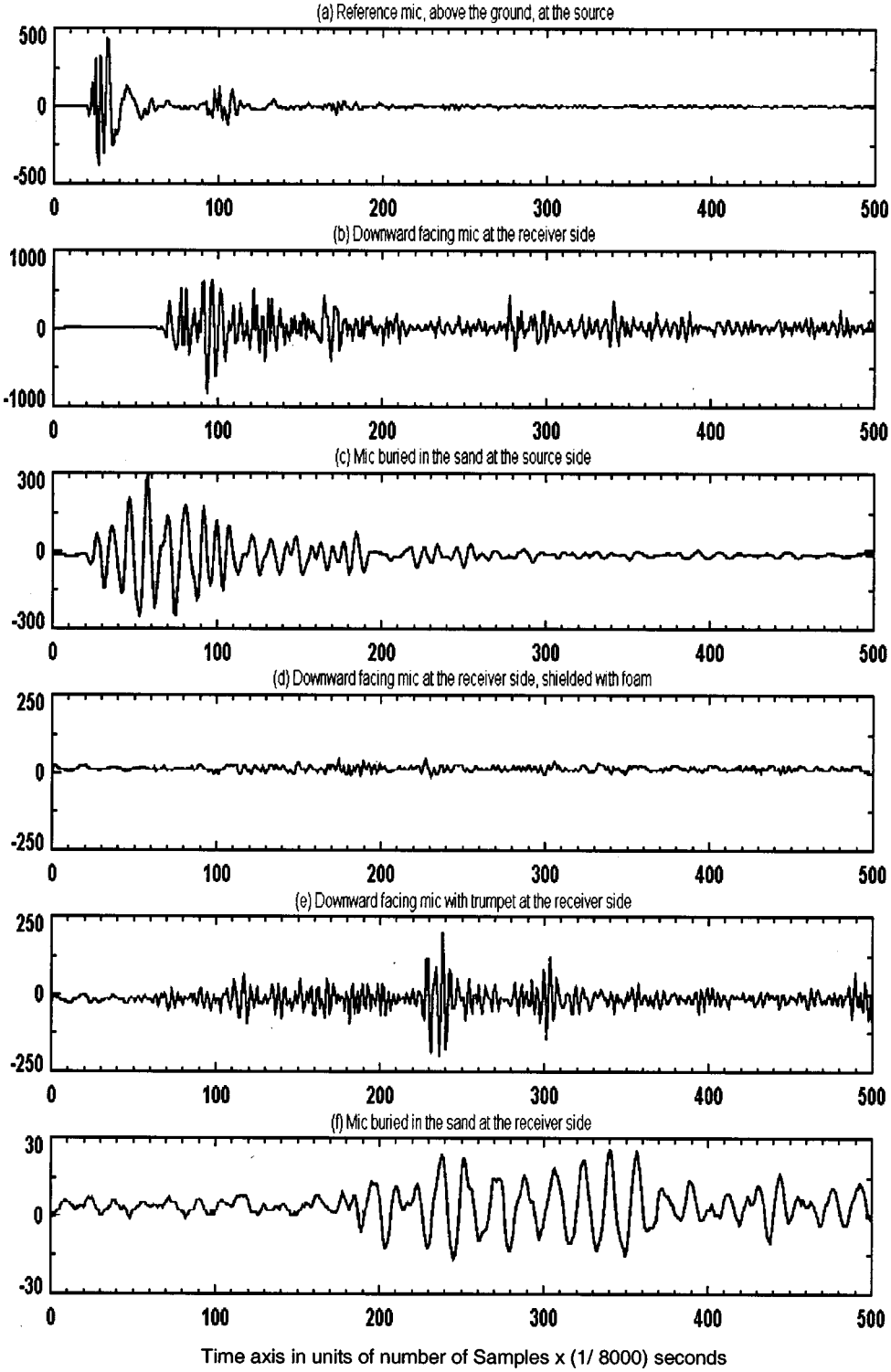


FIG. 3. Received time series for microphone arrangements shown in Fig. 2. (a), (b), (d), and (e) show received signals by microphone above the ground while (c) and (f) shows received signal by microphones buried in the sand. (a) and (c) are signals at acoustic source while (b), (d), (e), and (f) are signals at detector side.

(c) and (f), shown in Fig. 2, are buried in the sand (5 cm deep), and just below the microphones (a) and (b), respectively. The time delay between the signal wavefront arrivals of microphones (a) and (b) and between (c) and (f) corresponds to the time taken by the signal to travel between the corresponding microphones through the air and in the sand, respectively. These time delays can be estimated from the results of Fig. 3 that shows the corresponding time series of the microphones (a)–(f) of Fig. 2.

To maximize the microphones receiving characteristics, a number of experiments were conducted to determine the optimum position configuration for the microphone array. For example, microphone (d), shown in Fig. 2, was acoustically isolated from the surrounding acoustic signals propagating through the air by using thick foam to cover the sand surface. In another case, the microphone (e), shown in Fig. 2, was set in a plastic trumpet to maximize its receiving directivity pattern of the acoustic signals arriving from the ground.

From these experiments it became obvious that the amplitude of the received signals by microphone (d) [e.g., Fig. 3(d)] were much smaller than the corresponding amplitudes

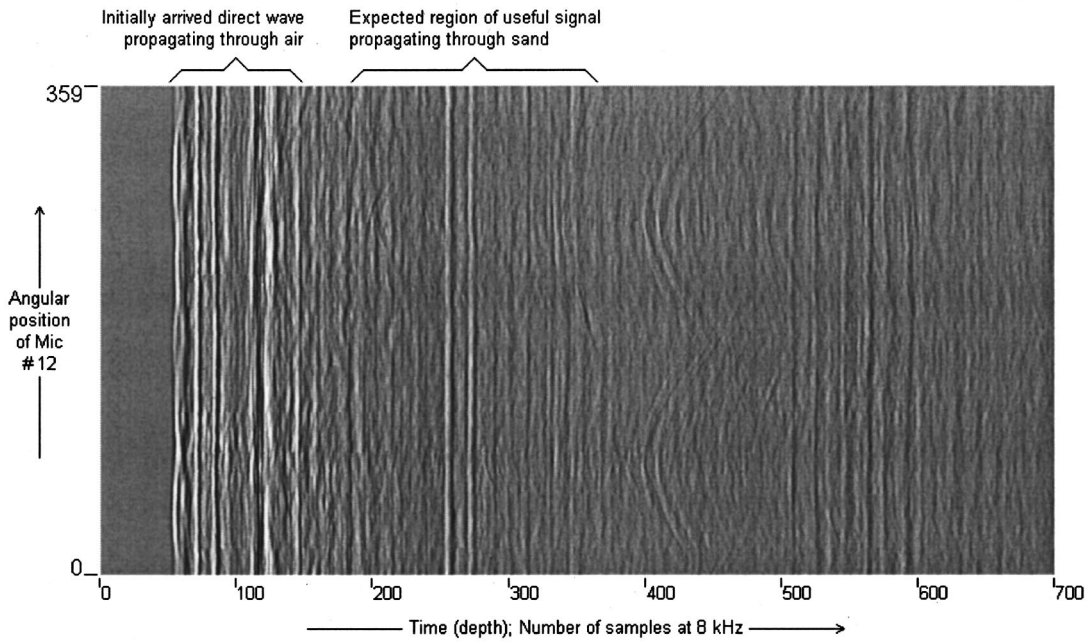


FIG. 4. Temporal alignment of received signals for the 12th microphone of the receiving array and for all the 360 projections. This signal display is called ''gram.''

of the signals of microphones (b) and (e), [e.g., Figs. 3(b), (e)]. In particular, the differences in amplitude between the received signals (b) and (e) in Fig. 3, show that a very significant portion of the energy collected by (b) is received directly from the air and hence the SNR of the component of the acoustic signal propagating through the ground is very small, as expected because of the very poor air-to-ground-to-air coupling. In another experimental set up, the amplitude of the acoustic signal received by microphone (e), [e.g., Fig. 3(e)] has smaller amplitude than that of the signal of microphone (b) [e.g., Fig. 3(b)]. This confirms that the dominant signal components of the received acoustic signals by the microphone (e) [e.g., Fig. 3(e)] are signals propagating in the ground. This observation is confirmed also by the signal of the shallow buried microphone (f), since the time arrival of the received signals for microphones (e) and (f) are nearly identical. Furthermore, it appears that the frequency spectrum for the times series (c) and (f) is different from those for the microphones in the air (d) and (e). This was expected since the ground acts as a low pass filter and the higher frequencies that do not penetrate the ground they are not included in (c) and (f).

Thus, the time delay of the signal wavefront arrivals between the reference microphone (a) [e.g., Fig. 3(a)] and microphones (e) and (f), [e.g., Figs. 3(e), (f)], provides estimates of the speed of the acoustic signals propagating through the sand, which are in the range of 80–100 meter/second, which agrees with a similar experiment reported in Ref. 25.

In summary, the segments of the received microphone time series in the range of 180–500 samples for the microphones (b), (e), and (f) include the signal components of the pulses propagating through the air-to-ground-to-air. We will use this segment of the received signal to define the sinograms for the purpose of underground image reconstruction. It is worth noting that although the signal loss due to the poor air-to-ground-to-air coupling was great, this signal loss was tolerable in the image reconstruction process. This was confirmed also from the image results that were reconstructed using these very low SNR signals. Another important observation is the increased duration of the received signal by microphone (f), shown in Fig. 3(f) as compared to that in Fig. 3(a). This is an indication that the received signal in Fig. 3(f) includes multipath propagation effects in the ground. This, however, cannot be considered as justification of our claim that time delay t_j in the projection data $p(r_n,\theta_i,t_j)$ can be equated with depth, since the experimental setup with the tank filled with sand is a nonstratified medium. Therefore, the reconstructed images of this investigation would represent near the surface 2D horizontal tomography sections of the underground medium.

C. Experimental results

Each experiment, which was based on the setup of Fig. 1, generated a three-dimensional data set of projections defined by

$$p(r_n,\theta_i,t_j) \quad (n=1,...,N), \quad (i=1,...,M), \quad (j=1,...,K), \tag{2}$$

where $N=32$ microphones (31 sensors in the receiving array plus the reference microphone), $M=360$ projections around a circle, and $K=8192$ samples for each microphone time series. These 360 sets of time series had to be aligned according to their reference microphone's signal [Fig. 3(a)]. Figure 4, shows the aligned received signals for the 12th (a randomly selected) microphone. Such a signal display is called ''gram.'' The alignment of the time series from the 360 different projections is clearly evident by the results of Fig. 4.

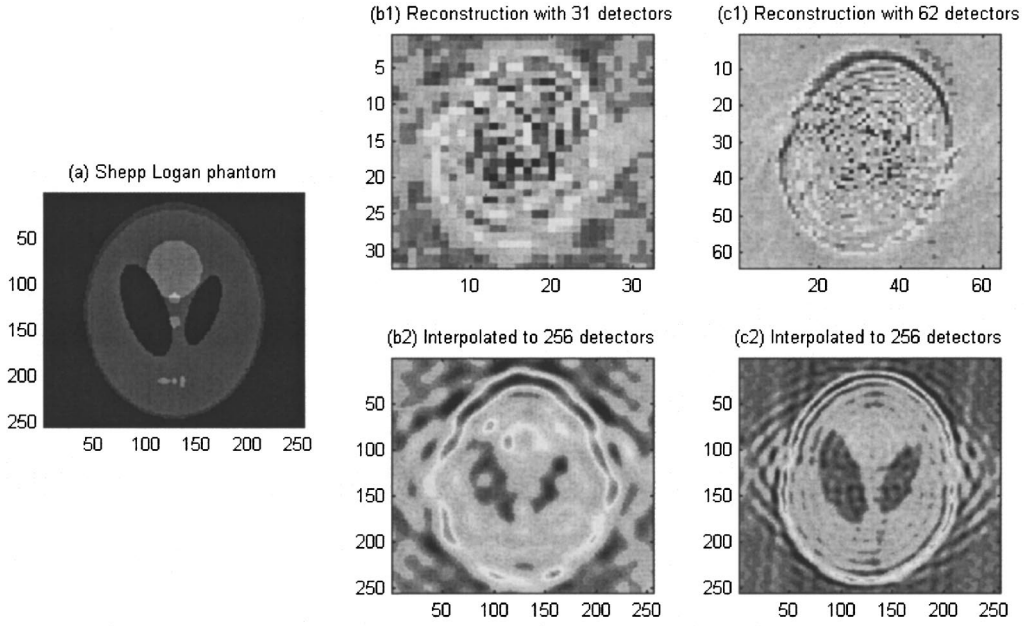


FIG. 5. Simulation results using a Shepp–Logan phantom for diffraction image reconstruction integrated with interpolation techniques. Left image shows the simulated Shepp–Logan phantom. Upper-middle shows the 256×256 reconstructed image from a sinogram ($N=31, M=360$), without interpolation. Lower-middle presents the 256×256 reconstructed image from the sinogram ($N=31, M=360$) that has been interpolated to ($N=256, M=360$). Upper-right shows the 256×256 reconstructed image from a sinogram ($N=62, M=360$), without interpolation. Lower-right presents the 256×256 reconstructed image from the sinogram ($N=62, M=360$) that has been interpolated to ($N=256, M=360$).

1. Interpolation for improved image resolution and simulations

The image resolution characteristics of a CT reconstructed image are related to the number of sensors N, the sensor spacing and the number of projections M [Eq. (2)]. A typical x-ray CT scanner deploys approximately 1400 sensors with 1 mm pitch and acquires 1100 projections to have sufficient image resolution for medical diagnostic purposes. In our case, however, due to limited resources, the present setup included only 31 sensors and the acquisition of 360 projections; hence the image resolution was anticipated to be very poor. Furthermore, because of our sensor spacing (e.g., pitch 3 cm) the spatial sampling frequency was 33 samples/meter, which indicated that the maximum spatial frequency that could be handled by this pitch was 16.5 samples/meter.[18] Another complication in terms of the image resolution capabilities of the current experimental setup was due to the fact that our experimental observations in Fig. 3 show that only acoustic frequencies below 1 kHz penetrated the sand medium [i.e., wavelength in the ground, (100 m/s)/(1000 Hz to 500 Hz) = 10 cm to 20 cm]. Thus, the image resolution was very limited in our diffraction tomography experiments since the wavelengths of the interrogating acoustic waves were in the range of 10 cm to 20 cm and the spatial sampling frequency was 33 samples/meter. As a result, spatial details higher than 16 sample/meter and object dimensions comparable to the wavelength of 10 cm to 20 cm cannot be detected with our current system configuration.

To minimize the impact of the poor image resolution inherent in our system, we used interpolation techniques, as defined in Ref. 18. To assess the performance of this interpolation technique we used diffraction tomography simulations for 800 Hz and the Shepp–Logan phantom[14,17,18] (left image in Fig. 5). The upper-middle image results of Fig. 5 have been generated from a sinogram that has the same parameters as in the experiment [i.e., $M=360$, $N=31$ (500–1000) Hz]. Image reconstruction results using the same sinogram and with interpolation parameters ($M=360$, $N=256$) are shown by the lower-middle image of Fig. 5. The right upper and lower reconstructed images of Fig. 5 indicate the anticipated improved image resolution performance of our experimental setup for $N=62$ and with smaller sensor spacing of 1.5 cm (e.g., pitch 1.5 cm). It is apparent from these simulations that the experimental configuration of $N=31$ sensors would not be sufficient to provide reasonable image resolution even for high SNR (which is the case in the simulations), since the wavelength of the interrogating acoustic pulses is comparable to the dimensions of the objects and the sensor spacing is not sufficiently small. Thus, for poor SNR, image resolution would be even worse. For more details about the image resolution limits of diffraction tomography in terms of the wavelength of the interrogating energy, the size of the object and the spatial sampling of the receiving sensor array, the reader may review Ref. 26.

2. Real results

Based on Eq. (2) and Fig. 4, a 2D sinogram for a predefined temporal sample is given by

$$p(r_n, \theta_i, t_{j=\text{predefined}}) \quad (n=1,...,N=31),\ (i=1,...,M=360),\ (j=\text{predefined}). \tag{3}$$

Figure 6(a) displays a sinogram for the time delay $j=376$. The correspondence between the data sets of Figs. 4 and 6(a) is that a vertical line for the sensor $n=12$ in Fig. 6(a) repre-

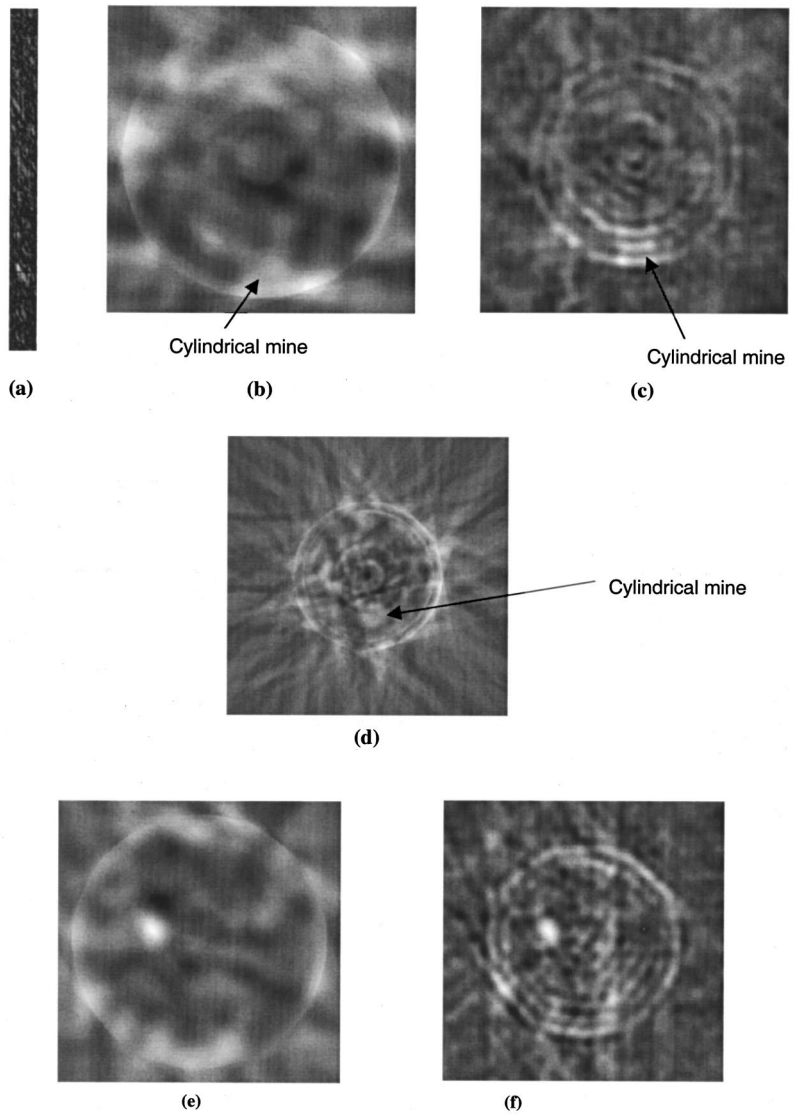


FIG. 6. (a) Sinogram from the data set of Fig. 4 corresponding to the temporal sample $j=376$, for the 31 microphones and 360 projections. (b) Image reconstruction based on filtered back projection and using the sinogram data set shown in (a) with no interpolation (wet sand). (c) Image reconstruction based on Fourier diffraction algorithm and using an interpolated version (31 to 256 sensors) of the same sinogram data set (a). The image of the buried mine is obvious at the bottom. This location coincides with the expected mine location in Fig. 1(b). (d) Image reconstruction based on filtered back projection of a data set from a previous experiment reported in Ref. 24. This image is for the same buried mine of Fig. 1(b). The mine location here is slightly different from (b) and (c). In this case the sand was dry. (e) Image reconstruction based on filtered back projection and using a sinogram data set with no interpolation (temporal sample $j=186$). In this experiment the sand was wet. (f) Image reconstruction based on Fourier diffraction algorithm and using an interpolated version of the sinogram data set from 31 to 256 sensors and for the same data set with those of image (e). The location of the white image spot in (e) and (f) coincides with that of the buried metallic object in Fig. 1(b).

sents the same data as the vertical line of Fig. 5 for the time delay $j=376$.

The image in Fig. 6(b) has been reconstructed from the sinogram of Fig. 6(a) and for the temporal sample $j=376$. It shows a circular pattern (indicated by the arrow) that its location coincides with the location and time delay depth of the buried plastic mine [Fig. 1(b)]. In this experiment the sand was wet in order to improve the acoustic propagation and compare results with dry sand. The image reconstruction algorithm for the results of Fig. 6(b) was the Fourier diffraction algorithm[18] using interpolated data (from 31 to 256 sensors) for the same temporal samples. Thus, the image in Fig. 6(c) has been derived from the same sinogram as the image in Fig. 6(b). Their differences represent better image resolution characteristics as those in the simulation results shown in Fig. 5.

As discussed earlier, the temporal samples for $j=180,...,500$, are expected to provide information about buried objects. Although the images of the buried mine (Fig. 6) have very poor image resolution, their spatial location coincides with that of the buried mine as in our previous experiments reported in Ref. 24 and shown by Fig. 6(d). The image in Fig. 6(c) was reconstructed using filtered back projection algorithm on noninterpolated data. The experiment was carried out in dry sand and for slightly different mine location and depth with those of Figs. 6(b) and (c). Figures 6(e) and (f) show reconstructed images for temporal sample $j=186$ using filter back projection and Fourier diffraction (interpolated data), respectively. The white spot in both figures corresponds to the expected location of the buried metallic mine, shown in Fig. 1(b).

D. Adaptive interference cancellation (AIC) in acoustic CT

The results in Fig. 7 demonstrate the advantages of using adaptive interference cancellation (AIC) processing to minimize the impact of the poor SNR in the image reconstruction process. The implementation of the AIC process in our acoustic CT concept has already been discussed briefly in Sec. II and for more details the reader may review the papers.[13,22,23] The AIC implementation effort requires that the 31-microphone array should be rearranged as follows: 16-microphones should be positioned along the receiving ar-

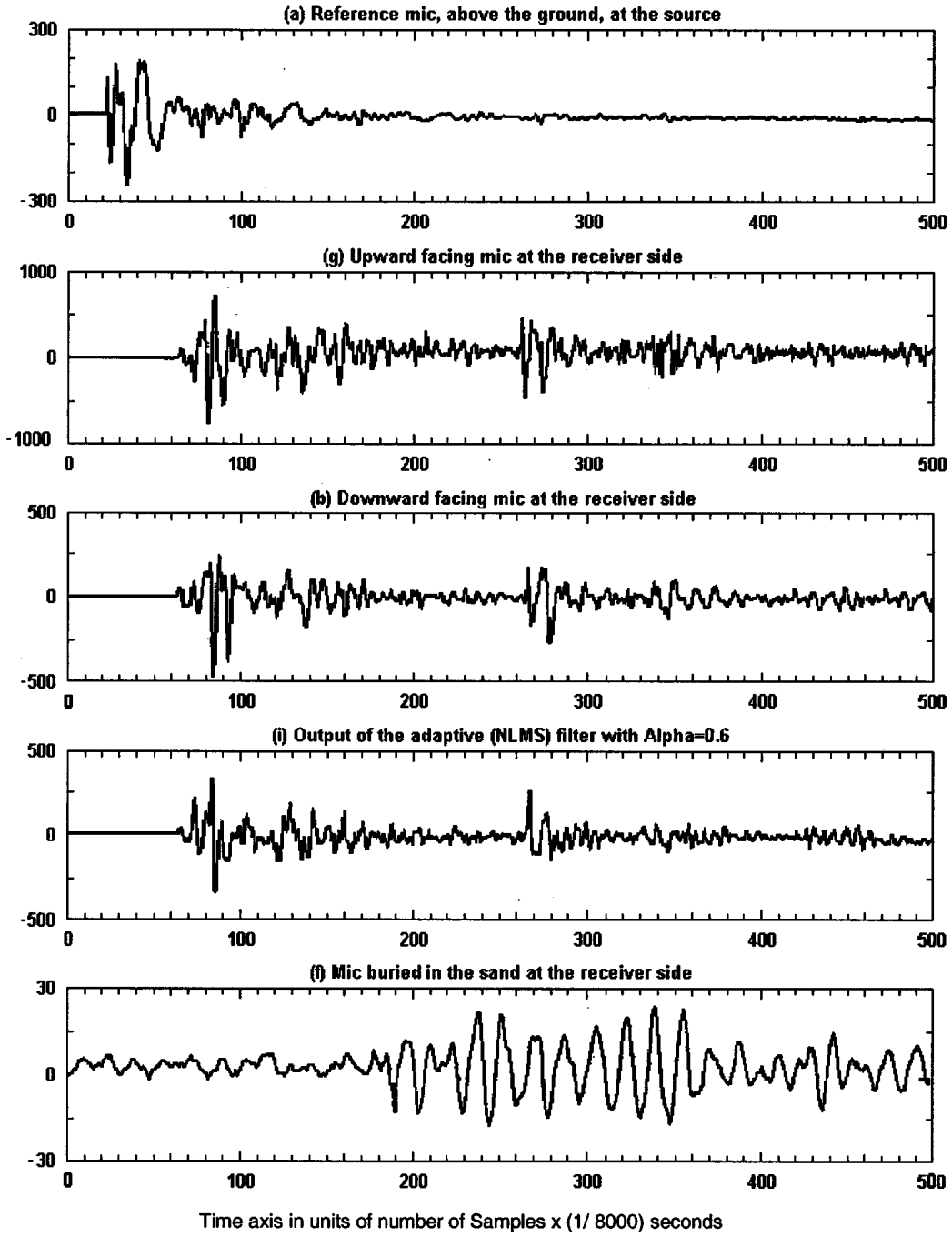


FIG. 7. Time series from the adaptive acoustic CT experiment. (a) is the signal from the reference microphone. (g) is the signal from the upward microphone that is considered as the noise (interference) in the AIC process. (b) is the signal from the downward microphone that is considered as the noisy signal in the AIC process. (i) is the output signal of the AIC process and (f) is the signal from the buried microphone.

ray downwards to receive the noise signal for the AIC process. 15-microphones should be positioned along the receiving array upwards to receive the interference, which is the propagating in the air component of the transmitted pulse. The spacing for both the upward and downward microphones was the same.

The time series in Fig. 7(a), refer to the reference microphone (a) (see Fig. 2), located near the array of the acoustic sources. Figures 7(g) and (b) show the received time series for the upward and the downward facing microphones that represent the interference and noisy signals, respectively, that are essential in an AIC process. The time series in Fig. 7(i) show the output of the AIC process and Fig. 7(f) shows the signal of the buried microphone that defines the time position of the wave front arrival in the ground. At this point it is important to note that a comparison between the time series (i) and (f) of Fig. 7 indicates that these two time series are not similar since AIC filtering process (because of the very low SNR) has not been very effective to remove all the components of air contributions in the received signal. However, the AIC process has been sufficient to enhance the image resolution for this experimental setup as shown by the image reconstruction results of Fig. 8.

Thus, for the temporal sample $j=55$ the image 8(a) and

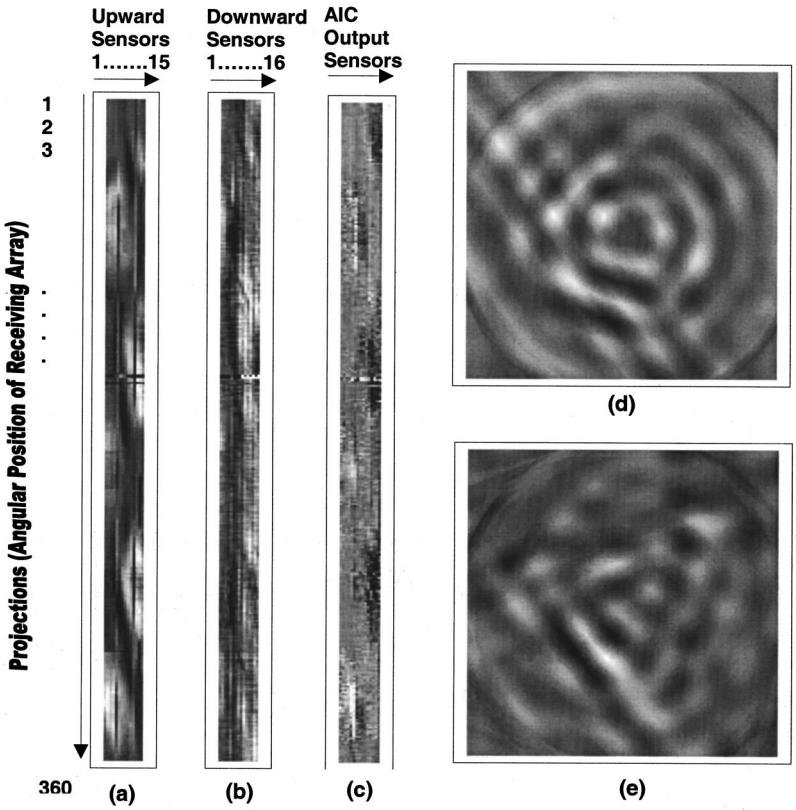


FIG. 8. Sinograms from the acoustic CT experiment and for the temporal sample $j=55$. (a) For the 15-upward microphones. (b) For the 16-downward microphones. (c) Output of the AIC process. (d) Image reconstruction using the sinogram shown in (b). (e) Image reconstruction using the AIC sinogram shown in (c). The AIC image output shows the expected triangular base, shown in Fig. 1.

(b) are the sinograms for the 15-upward (interference) and the 16-downward (noisy signal) microphones, respectively. These sinograms define the inputs for the AIC processor. The output of the AIC processor is the sinogram shown in Fig. 8(c), which is supposed to have improved SNR as compared with the SNR of Fig. 8(b).

The image reconstructed from the AIC output sinogram of Fig. 8(c), is shown in Fig. 8(e). This image result [Fig. 8(e)] demonstrate the efficiency of the AIC processor in minimizing the impact of the air-reverberations in the downward microphones, when compared with the image of Fig. 8(d), which has been derived from the image reconstruction of the noisy sinogram of Fig. 8(b) of the downward facing microphones.

The features in Fig. 8(e) represent objects and structures above the ground (the triangular base shown in Fig. 1), as this was expected since $j=55$, which indicates the early arrival of the signal wave fronts propagating through the air. It is interesting to note that these triangular base features were completely masked in the image results of Fig. 8(d), and were only revealed by the AIC process in Fig. 8(e). Although, the number of sensors in Fig. 8(e) were halved from those in Fig. 6(b), the AIC process appeared to be effective in improving image quality by minimizing the impact of the surrounding interference noise. However, since the microphone spacing in the AIC process was 6 cm, the spatial sampling frequency was 16 samples/meter, which indicates that the maximum spatial frequency that could be handled by this pitch would be insufficient to sample the dimensions of the buried mines, as indicated also by the simulation results of Fig. 5. Therefore, because of the limited number of microphones that were available in this experimental study, we were unable to assess the effectiveness of the AIC process in providing better image resolution for the mines shown in Fig. 6.

E. Key issues and future directions

The results of this experimental study and the discussions in the preceding sections suggest that a successful implementation of the proposed acoustic tomography concept as a nondestructive underground imaging process should address the following issues.

The receiving microphone array should have at least 64-downward-microphone (preferably 128-microphones) with smaller array pitch to improve the image resolution.

The AIC process is essential to minimize the effects of the signal arrivals propagating through the air. Thus, the number of sensors should be increased to 128 (preferably 256) to include the 64 (preferably 128) upward microphones.

The pulse design characteristics should guide the design efforts for the acoustic source array to avoid the nonlinearities discussed in Sec. IV A and to include a frequency range that would allow ground penetration with sufficiently small wavelengths for improved image resolution.

Finally, if the AIC process would be effective to improve the SNR, the replica correlation process should be implemented on the microphone time series in order to differentiate the signal arrivals from various underground structures and from the air.

Although, the available resources were very limited and not sufficient to address the above issues, the results discussed in Sec. IV, suggest that the proposed acoustic tomog-

raphy concept can be used successfully as a nondestructive underground imaging technique.

At this point it is important to note that the poor image results of our investigation are in agreement with the findings of another investigation by Crawford and Kak[20] that proposed a technique for breast tomography imaging using ultrasonic CT concepts. Although their physical parameters (forward problem) were more favorable than the poor air-to-ground-to-air coupling (and multipath) imposed in our experimental design, the image quality of their experimental results was very poor and similar with those reported in our paper. In summary, investigators should anticipate that because of the very low frequencies (e.g., long wavelengths) used in acoustic CT imaging applications the associated image resolution would be poor and that they should not expect the high image resolution that is available by the x-ray CT scanners.

V. CONCLUSION

The present paper describes a new approach to nondestructive subsurface imaging for detecting landmines and other buried objects. Subsurface imaging is performed by sending acoustic energy into the soil around a circular pattern and collected at the diametrically opposite end of it by a linear microphone array. From these collected signals, subsurface images are reconstructed employing image reconstruction algorithms, in the same manner as done in CT imaging. It is assumed that acoustic waves that propagate in the ground carry information about the subsurface buried objects to the sensors of the receiving array. It is also assumed that in the received signal, the signal amplitude at different time delays carries information from the different depths of the stratified soil. Consequently, we would be able to reconstruct the cross sectional images at different underground depths.

Among several challenges, the most important was the limited knowledge of the propagation characteristics. The very weak signals of interest, received by the microphones, resulted in poor SNR that was another major problem due to the poor air-to-ground-to-air acoustic coupling.

In summary, the results of this investigation provide supporting arguments that the implementation of the AIC process in combination with the deployment of large number of sensors in the receiving array could be a valuable method for 3D subsurface imaging for the detection of shallow buried objects using acoustic diffraction CT.

[1] K. A. Dines and J. Lytle, ''Computerized geophysical tomography,'' Proc. IEEE **67**, 1065–1073 (1979).

[2] W. C. King, A. J. Witten, and G. D. Reed, ''Detection and imaging of buried wastes using seismic wave propagation,'' ASCE J. Environ. Eng. **115**, 527–540 (1989).

[3] A. J. Devaney, ''Geophysical diffraction tomography,'' IEEE Trans. Geosci. Remote Sens. **GE-22**, 3–13 (1984).

[4] A. J. Witten and E. Long, ''Shallow applications of geophysical diffraction tomography,'' IEEE Trans. Geosci. Remote Sens. **GE-24**, 654–662 (1986).

[5] Reporting on Science and Technology, ''Archaeology—Underground Noises,'' The Economist, pp. 83, issue of July 2000.

[6] S. Stergiopoulos, ''Advanced beamformers,'' *Advanced Signal Processing Handbook: Theory and Implementation for Radar, Sonar and Medical Imaging Systems* (CRC, Boca Raton, 2000), Chap. 6.

[7] G. Sakas, G. Karangelis, and A. Pommert, ''Advanced applications of volume visualization methods in medicine,'' *Advanced Signal Processing Handbook: Theory and Implementation for Radar, Sonar and Medical Imaging Systems* (CRC, Boca Raton, 2000), Chap. 7.

[8] J. Haystead, ''Portable systems apply infrared technology to detect and classify mines,'' Vision Systems Design, 36–40, 1997.

[9] C. J. Vaughan, ''Ground penetrating radar survey in archeological investigations,'' Geophysics **51**, 595–604 (1986).

[10] J. McFee *et al.*, ''CRAD Countermine R&D Study—Final Report,'' DRES-SSP-174, Defense Research Establishment Suffield, Rlaston, Alberta, Canada, 1994.

[11] R. C. Benson, R. P. Glaccum, and M. R. Noel, *Geophysical Techniques for Sensing Buried Wastes and Waste Migration* (U.S. Envir. Prot. Agency, Washington, DC, 1982).

[12] C. H. Frazier, N. Cadalli, D. C. Munson, and W. D. O'Brien, ''Acoustic imaging of objects buried in soil,'' J. Acoust. Soc. Am. **108**, 147–156 (2000).

[13] A. Dhanantwari, S. Stergiopoulos, and I. Iakovides, ''Correcting organ motion artifacts in x-ray CT medical imaging systems by adaptive processing (Part I: Theory),'' Med. Phys. **28**, 1562–1576 (2001).

[14] A. C. Kak, ''Computerized tomography with x-ray, emission, and ultrasound sources,'' Proc. IEEE **67**, 1245–1272 (1979).

[15] J. Radon, ''Uber die Bestimmung von Funktionen durch ihre Integral werte langs gewisser Mannigfaltigketiten,'' Ber. Verh. Saechs. Akad. Wisse. Leipzig, Math.-Phys. Kl. **69**, 262–279 (1917).

[16] A. J. Devaney, ''A filtered back projection algorithm for diffraction tomography,'' Ultrason. Imaging **4**, 336–350 (1982).

[17] A. C. Kak, *Digital Image Processing Techniques* (Academic, New York, 1984).

[18] S. X. Pan and A. C. Kak, ''A computational study of reconstruction algorithms for diffraction tomography: Interpolation versus filtered backpropagation,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-31**, 1262–1275 (1983).

[19] R. K. Mueller, M. Kaveh, and G. Wade, ''Reconstructive tomography and applications to ultrasonics,'' Proc. IEEE **67**, 567–587 (1979).

[20] C. R. Crawford and A. C. Kak, ''Multipath artifact corrections in ultrasonic transmission tomography,'' Ultrason. Imaging **4**, 234–266 (1982).

[21] K. Attenborough *et al.*, ''The acoustic transfer function at the surface of a layered poroelastic soil,'' J. Acoust. Soc. Am. **79**, 1353–1358 (1986).

[22] S. Stergiopoulos, ''Implementation of adaptive and synthetic aperture beamformers in sonar systems,'' Proc. IEEE **86**, 358–396 (1998).

[23] B. Widrow and M. E. Hoff, Jr., ''Adaptive noise cancellation: Principles and applications,'' Proc. IEEE **63**, 1692–1716 (1975).

[24] S. Stergiopoulos, R. Alterson, D. Havelock, and J. Grodski, ''Acoustic 3D computed tomography for demining and archeological applications,'' J. Acoust. Soc. Am. **107**, 2806 (2000).

[25] G. D. Larson *et al.*, ''Air acoustic sensing of seismic waves,'' J. Acoust. Soc. Am. **107**, 2896 (2000).

[26] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE Press, 1988 or Society of Industrial and Applied Mathematics, 2001).

A combination of PZT and EMAT transducers for interface inspection

Won-Bae Na
Arizona Department of Transportation, Phoenix, Arizona 85007

Tribikram Kundu
Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona 85721

(Received 23 May 2001; accepted for publication 21 February 2002)

A PZT (Lead Zirconate–Titanate) transducer requires a couplant to send and receive mechanical waves. This requirement is a major shortcoming of the PZT technique for use in field applications. In the laboratory environment careful considerations and surface treatments are required to use PZT because the couplant can affect the consistency of experimental results. One alternative to overcome this shortcoming is the use of EMAT (ElectroMagnetic Acoustic Transducer). However, EMAT gives relatively low transmitted ultrasonic energy, with low signal to noise ratio, and the induced energy is critically dependent on the probe proximity to the test object. These are not desirable properties for NDT (nondestructive testing) of civil infrastructures. That is why, in this paper, a combination of PZT and EMAT is introduced for investigating reinforced concrete structures. Interface defects between steel bars and concrete are investigated by this technique. It is shown that the PZT–EMAT combination is very effective for steel bar–concrete interface inspection and the guided waves are useful for nondestructive testing of civil infrastructures. [DOI: 10.1121/1.1470503]

PACS numbers: 43.58.Vb, 43.35.Zc, 43.38.Dv, 43.38.Fx [SLE]

I. INTRODUCTION

The main limitation of a typical PZT (Lead Zirconate–Titanate) transducer for field applications is the need of using a couplant such as water, Vaseline, oil, honey, and so on. It is difficult to hold the couplants, especially water, in most field applications. In addition, these couplants may affect the consistency of experimental results. This problem led to the development of various fixtures including water-filled boots and wheels, bubblers, and squirters. Patton and Hsu[1] developed a dripless bubbler where water is recycled using a vacuum pump. Guo and Kundu[2–5] developed transducer holder mechanisms for pipeline inspection using solid couplers or liquid couplers in conical containers. This problem also led to the development of alternative dry-coupling techniques including the use of air-coupled piezoelectric transducers, laser ultrasonics, magnetostrictive transducers, and electromagnetic transducers.[6,7]

Air-coupled transducers are one solution to overcome the coupling problem.[8] To enhance the transmitted energy, the transducers need to be used with no backing layers, thus taking advantage of the high mechanical quality factor of piezoelectric disks. In addition, to improve the generation and reception efficiencies of the transducer, its front protection layer is made of a thin, porous material having low specific acoustic impedance.

Laser ultrasonics is another effective method of inducing and receiving ultrasonic waves without the need for a couplant.[9] The received signals are evaluated in a manner similar to the pulse–echo technique. However, laser ultrasonics is not very promising for inspecting civil infrastructures or in field applications, although it has application potentials in aerospace and mechanical industries in laboratory environment.

A magnetostrictive transducer uses the magnetostrictive and the inverse magnetostrictive effect. In the past, the development of magnetostrictive devices declined because the energy conversion efficiency of the magnetostrictive transducer is generally low. However, this transducer has a good application potential at relatively low frequencies and doesn't need any couplant.[7] Some investigators have developed a new magnetostrictive device by making use of the magnetostrictive properties of the material under insulation.[10] This device works efficiently on ferromagnetic materials and can be made to work on nonferromagnetic metals and even nonmetals, by providing a ferromagnetic material where sensors are to be placed. This transducer has been used to inspect strands, pipes, and rods.

EMAT (ElectroMagnetic Acoustic Transducer) generates ultrasound waves in electrically conductive materials by Lorentz force, electromagnetostrictive effect, or a combination of both. Both phenomena affect the atomic grid of the specimen so that acoustic waves are generated directly within the test specimen, instead of inside the transducers as in the case of PZT. Thus, it is a noncontact device.[11] The main advantage of EMAT over conventional PZT is that it does not need any couplant;[12,13] thus, we can eliminate the inconsistency arising from the couplant use during the nondestructive inspection. This characteristic attracts investigators to EMAT for determining material properties or detecting defects in structures. Another major advantage of this device is that different guided modes can be generated by simply changing the coil or magnet geometry, and their sequence of excitation. Thus, the transducer can induce specific ultrasonic modes including normal beam and angle-beam shear waves,

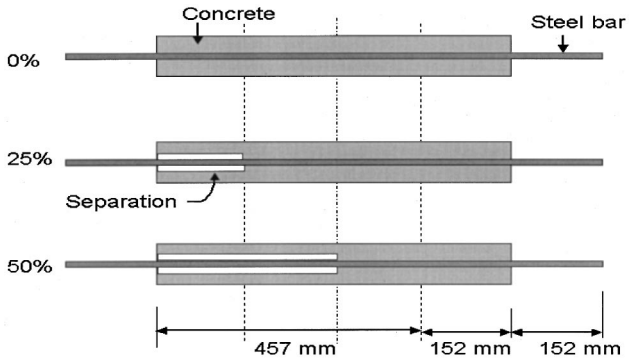


FIG. 1. The specimen inspected.

horizontal shear waves, Rayleigh waves, and Lamb waves.[14] This helps investigators to easily select a transducer for a specific test. The other advantages of EMAT are: (1) it can efficiently inspect austenitic welds by using horizontally polarized shear waves,[15] and (2) it can operate at high temperatures.[16] However, this transducer also has disadvantages: (1) this transducer can be used only for ferrous or nonferrous metals and cannot be used for inspecting nonmetallic materials such as FRP (fiber reinforced polymer); (2) it gives a relatively low transmitted ultrasonic energy and a low signal to noise ratio; and (3) the induced energy is critically dependent on the probe proximity to the test object, which for practical applications is commonly maintained below 1 mm.

In this research, the PZT–EMAT combination is used to generate and receive cylindrical guided waves in steel bars embedded in concrete. Cylindrical guided (or Lamb) waves are guided waves propagating in cylindrical structures such as bars and pipes. Important characteristics of the guided waves are that they can propagate a long distance and have several wave modes that are sensitive to different types of defects. Thus, in comparison to conventional ultrasonic techniques that use reflection, transmission, and scattering of longitudinal or shear waves by internal defects, the techniques employing cylindrical guided waves have a great potential for use in the nondestructive testing (NDT) of civil infrastructures. Because of its potential, some investigators have applied the guided wave technique to inspect interface defects in civil infrastructures. Pavlakovic *et al.*[17] used time-of-flight differences to identify voids at the interface between grout and steel tendons. Jung *et al.*[18] used a Lamb wave technique for detecting anomalies or discontinuities in concrete beams. Na *et al.*[19,20] used an inclination angle and frequency sweep technique for detecting interface delamination between a steel bar and concrete. It has been shown that this technique can be used efficiently for inspecting the degree of interface separation if guided waves are used.[19,20]

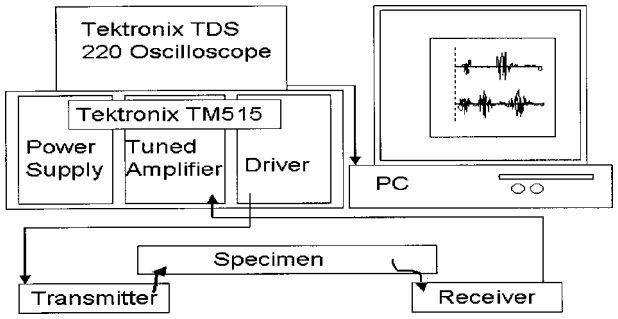


FIG. 2. The experimental setup using piezoelectric and electromagnetic acoustic transducers.

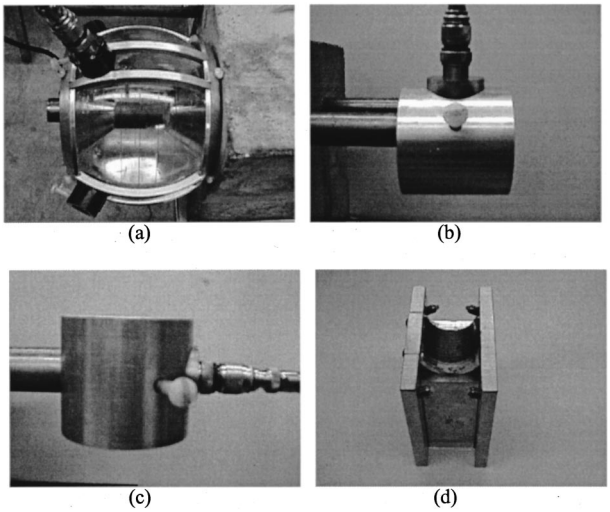


FIG. 3. Photographs of transducer holders: (a) spherical solid-coupler holder; (b) holder for PZT in direct contact at the surface of the steel bar; (c) holder for PZT in direct contact at the end of the steel bar; and (d) EMAT.

All of the above applications have used PZT transducers for exciting, propagating, and receiving guided waves. In these cases, careful considerations are required because any inconsistency in the couplant use or surface finish can affect

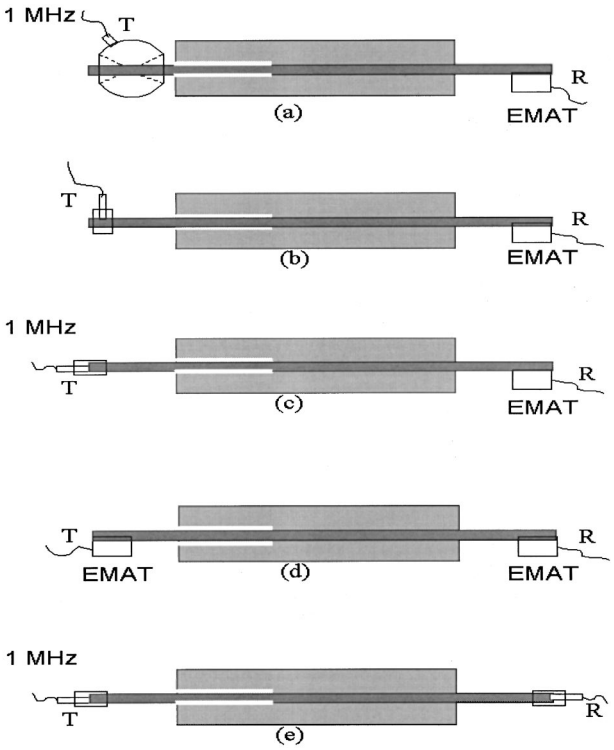


FIG. 4. Different transmitter–receiver arrangements using piezoelectric and electromagnetic acoustic transducers: (a) inclined PZT and EMAT; (b) PZT perpendicular to the axis of the steel bar and EMAT; (c) PZT oriented along the axis of the steel bar and EMAT; (d) EMAT and EMAT; and (e) PZTs oriented along the axis of the steel bar.

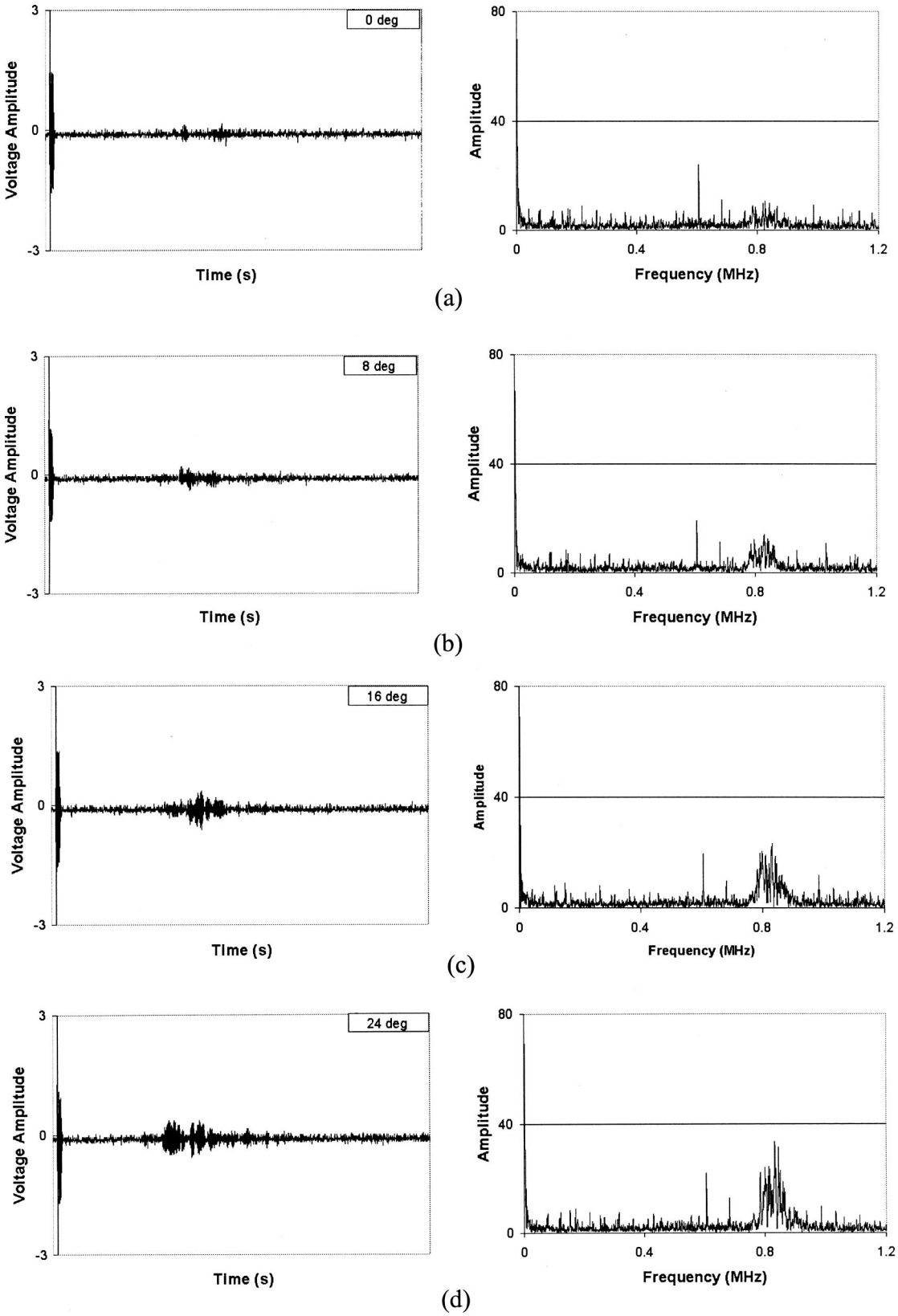


FIG. 5. $V(t)$ and $V(f)$ curves using the transmitter–receiver arrangement, shown in Fig. 1(a) for different incident angles: (a) 0°; (b) 8°; (c) 16°; and (d) 24°.

the consistency of the experimental results. Thus, in general, multiple tests are necessary and recommended for ultrasonic testing with PZT. It restricts civil engineers' ability to inspect several regions of civil structures with ultrasonic guided waves. This is why EMAT appears to be an attractive alternative to replace the conventional PZT. However, EMAT also has shortcomings for inspecting civil infrastructures, since it gives relatively low transmitted ultrasonic energy and low signal to noise ratio, and the induced energy is critically dependent on the probe proximity to the test object. These

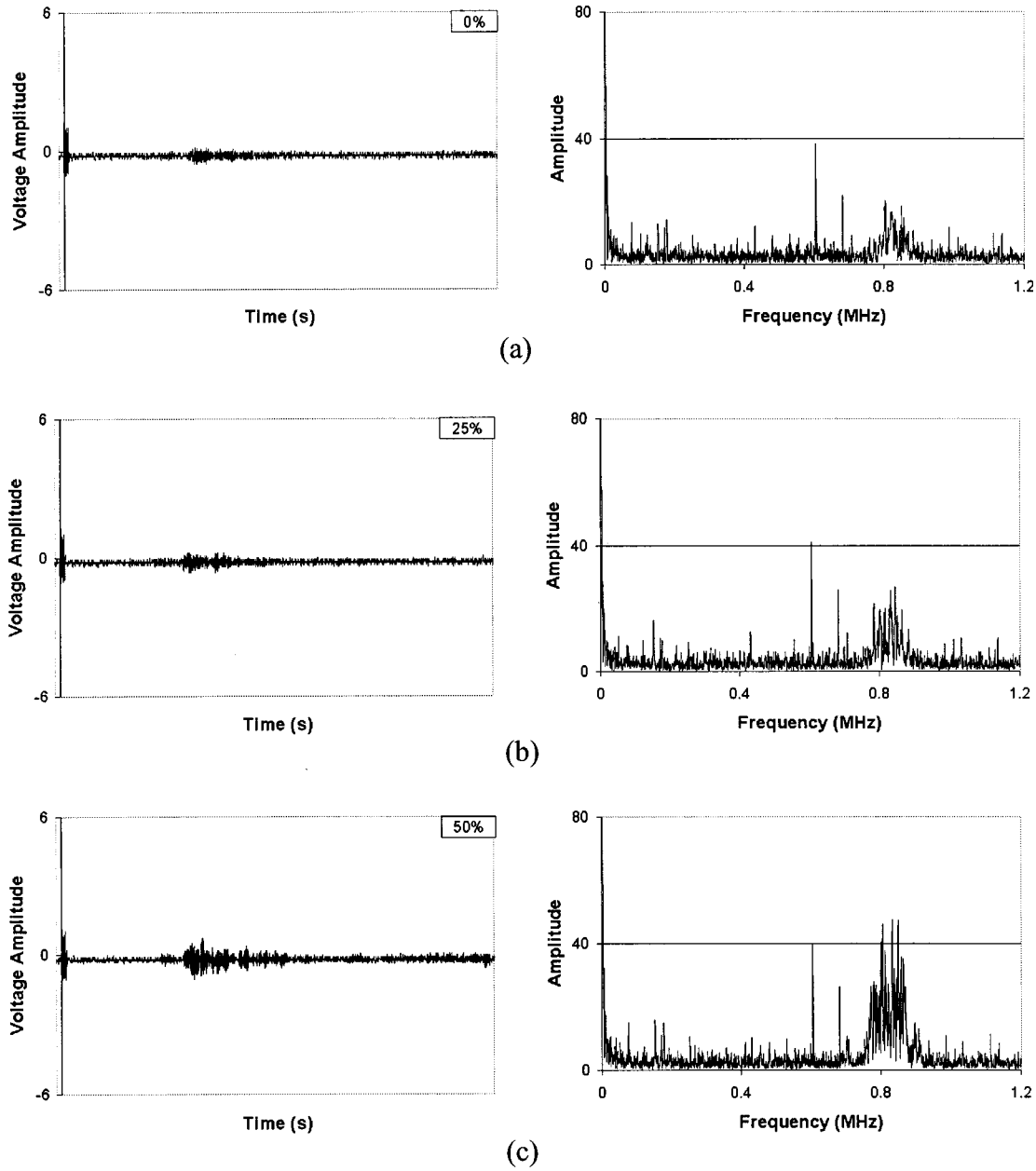


FIG. 6. $V(t)$ and $V(f)$ curves using the transmitter–receiver arrangement, shown in Fig. 4(a) for: (a) 0%, (b) 25%, and (c) 50% delaminated specimens. The transmitter inclination angle is 24°.

are not desirable properties for NDT of civil infrastructures. In this paper, a combined use of PZT and EMAT is introduced for inspecting civil infrastructures. It is hoped that the combination would minimize the disadvantages of both these techniques.

The goal of this paper is to investigate the feasibility of the combined use of EMAT and PZT for inspecting civil infrastructure. The specific defect studied here is an interface defect, a delamination between a steel bar and concrete. The feasibility of detecting the interface defect is investigated. For this purpose, three specimens with three different degrees of separation are fabricated. Each of these specimens contains a steel bar with different amounts of separation, 0%, 25%, and 50%, of the concrete–steel bar interface. These separations are fabricated artificially to idealize an interface delamination. Several transmitter–receiver arrangements combining PZT and EMAT have been devised to generate, propagate, and receive cylindrical guided wave through the specimens. From the tests, time history curves or $V(t)$ curves [voltage versus time] are obtained, then $V(t)$ curves are converted to $V(f)$ curves (voltage versus frequency), received signal amplitude versus frequency curves, by fast Fourier transform (FFT). These curves are used to quantify the degree of delamination of each specimen. This study shows one possible way to utilize EMAT for inspecting civil infrastructures.

II. EXPERIMENTAL SETUP

Three specimens were fabricated and tested. Non-air-entrained concrete and mild steel bars were used for the specimens. The weight proportion of cement: fine aggregate:

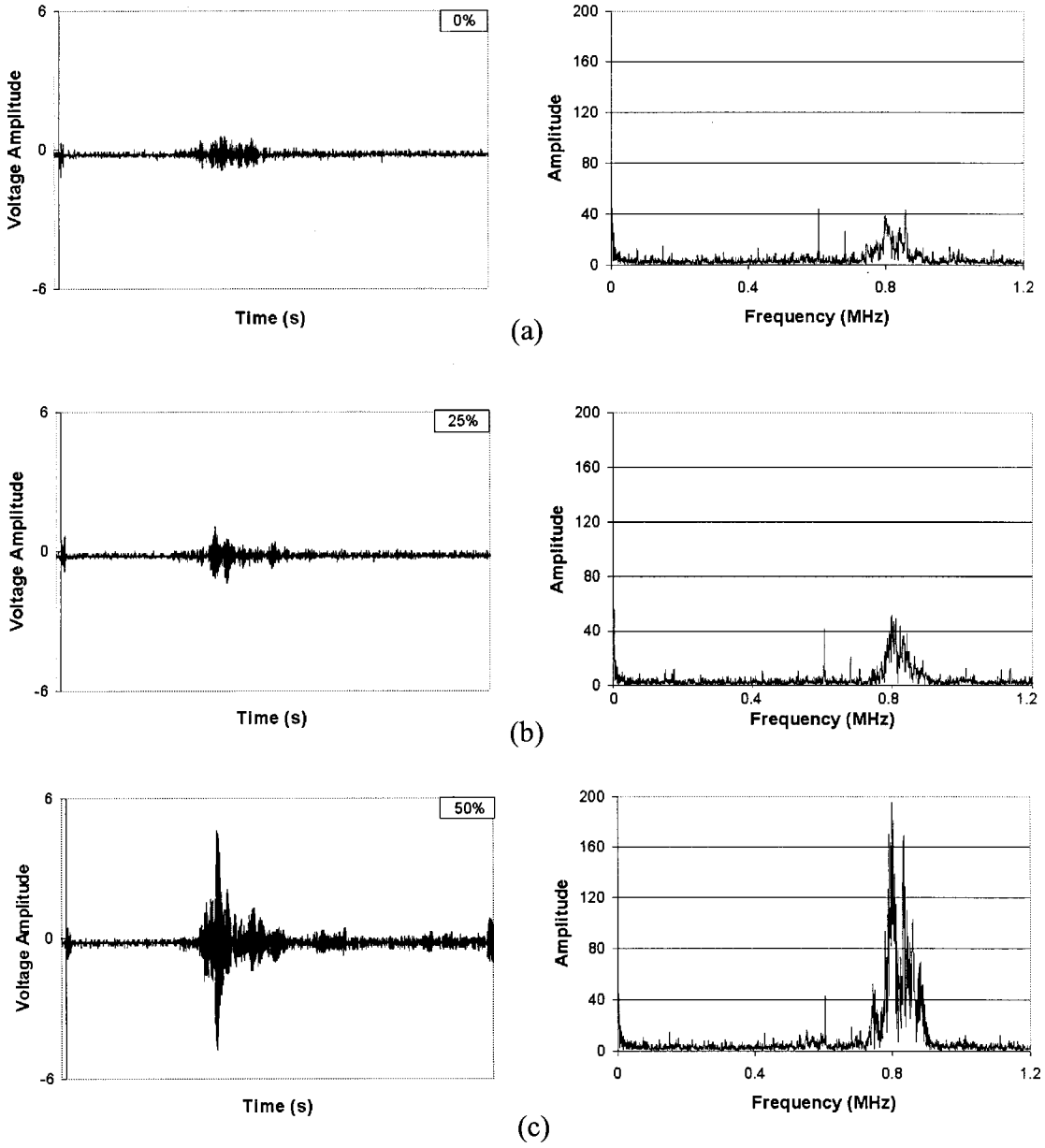


FIG. 7. $V(t)$ and $V(f)$ curves using the transmitter–receiver arrangement shown in Fig. 4(b): (a) 0%; (b) 25%; and (c) 50% delaminated specimens.

coarse aggregate: water used to make 0.028 cubic meter of concrete was 13.17: 26.79: 20.88: 4.99 kg. The concrete was mixed well using a mixer and a vibrator. Before testing, the specimens were left in the laboratory at room temperature for about $2\frac{1}{2}$ years. The steel bars were rust-free.

Geometries of the three specimens are shown in Fig. 1. The cross-sectional dimension of the specimens is 127 $\times$127 mm and the diameter of steel bar is 22 mm. To artificially make the delaminated regions, PVC pipes were placed around the steel bar and then extracted at 6 h after placing the concrete into wooden molds. Thus, the delaminated specimens have 0% (perfectly bonded), 25%, and 50% delaminated regions, respectively. For these specimens detachments or delaminations are open to the left side. Close detachments, not studied here, should produce similar results.

An experimental setup was designed as shown in Fig. 2. The transmitter was activated by a Tektronix TM 515 driver and the received signal was amplified by a Tektronix TM 505 amplifier and then displayed on a Tektronix TDS 200 oscilloscope as a function of time. The curves produced in this manner are denoted as $V(t)$ curves. Finally, the $V(t)$ curves were recorded and converted to $V(f)$ curves by taking fast Fourier transform.

A PZT transducer, manufactured by Harrisonic, with 1-MHz central frequency and an EMAT, manufactured by Industrial Sensors and Actuators, with 3.5-mm wavelength, were used as the transmitter as well as the receiver, and the received signals were compared. These transducers are shown in Fig. 3. This figure also shows several transducer holders for holding the transducers and generating cylindrical guided wave modes through the specimens. Figure 3(a) shows a spherical solid-coupler transducer holder for propagating antisymmetric (or flexural) guided wave modes with

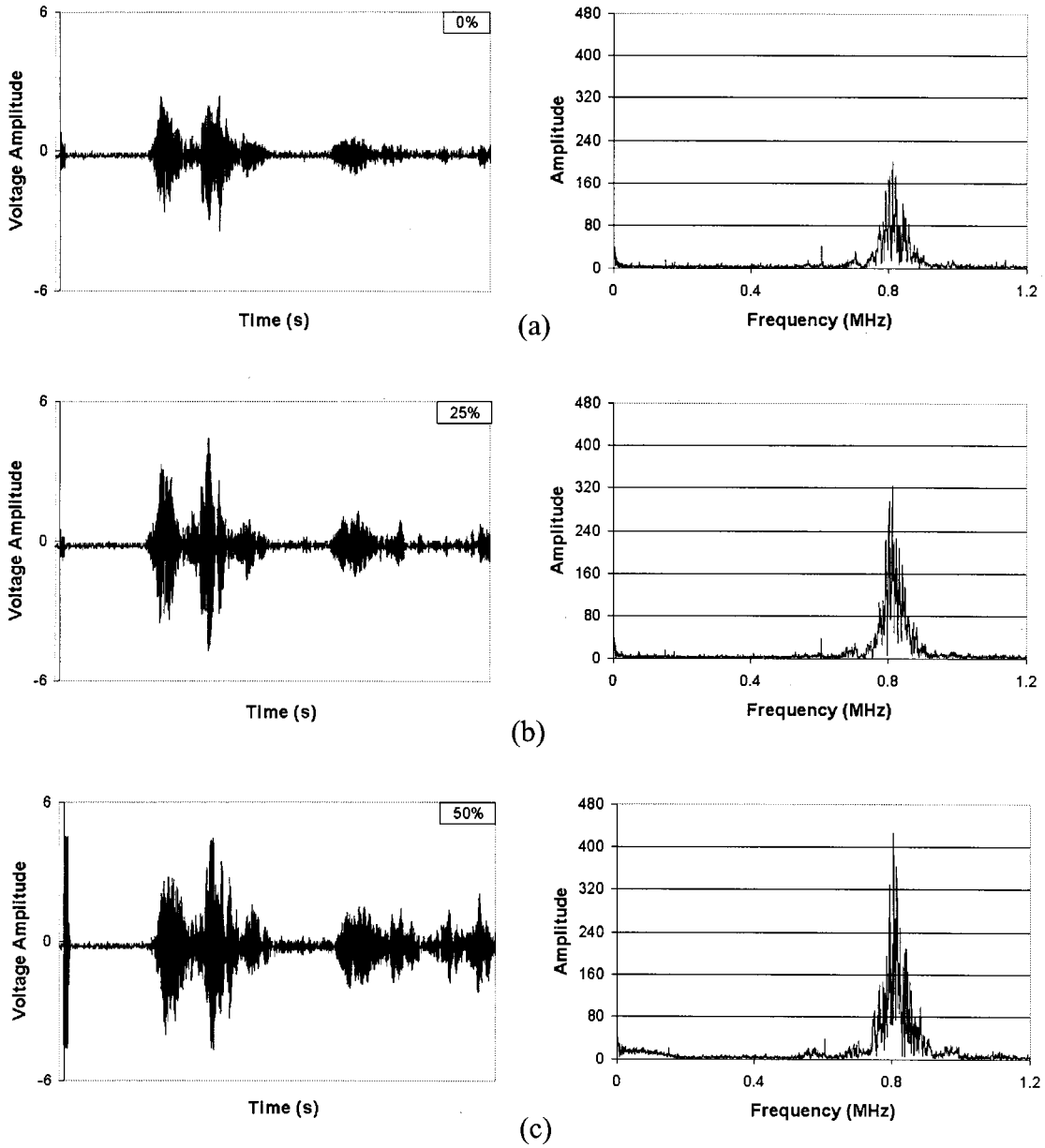


FIG. 8. $V(t)$ and $V(f)$ curves using the transmitter–receiver arrangement shown in Fig. 4(c): (a) 0%; (b) 25%; and (c) 50% delaminated specimens.

one transmitter. One can change incident angle from 0° to 27° with this holder. Figure 3(b) also shows a holder for propagating antisymmetric (or flexural) guided wave modes through the steel bar. This holder keeps the transmitter in direct contact with the circumferential surface of the steel bar. Figure 3(c) shows a holder for propagating longitudinal guided wave modes through the bar. This holder also holds the transmitter in direct contact with the steel bar at its end. Figure 3(d) shows an EMAT capable of propagating antisymmetric guided wave modes. In this case, we did not need any transducer holder since EMAT itself has a holding mechanism. Vaseline was used as the coupling medium for PZT.

With the transducers and the transducer holders as described in Fig. 3, several transmitter–receiver arrangements were tried out. Figure 4 shows different transmitter–receiver arrangements for which experimental results are presented. For all arrangements shown in Fig. 4 except the arrangement of Fig. 4(e), EMAT was used as the receiver and the previously described transducer holders and PZT (in one case EMAT) were used for generating cylindrical guided wave modes through the specimens. For the arrangement of Fig. 4(e), PZT transducers were used for generating as well as receiving longitudinal cylindrical guided wave modes through the specimen.

III. EXPERIMENTAL RESULT

Only one test result is shown for each case. However, multiple tests were carried out and the test results were found to be quite consistent when the transmitter and receiver positions and the coupling conditions were kept unchanged. Figure 5 shows $V(t)$ and $V(f)$ curves of the 25% delaminated specimen using the transmitter–receiver arrangement shown in Fig. 4(a) for different incident angles, 0°, 8°, 16°, and 24°. These curves show that the 24° incident angle gives the strongest signal. This incident angle was selected for

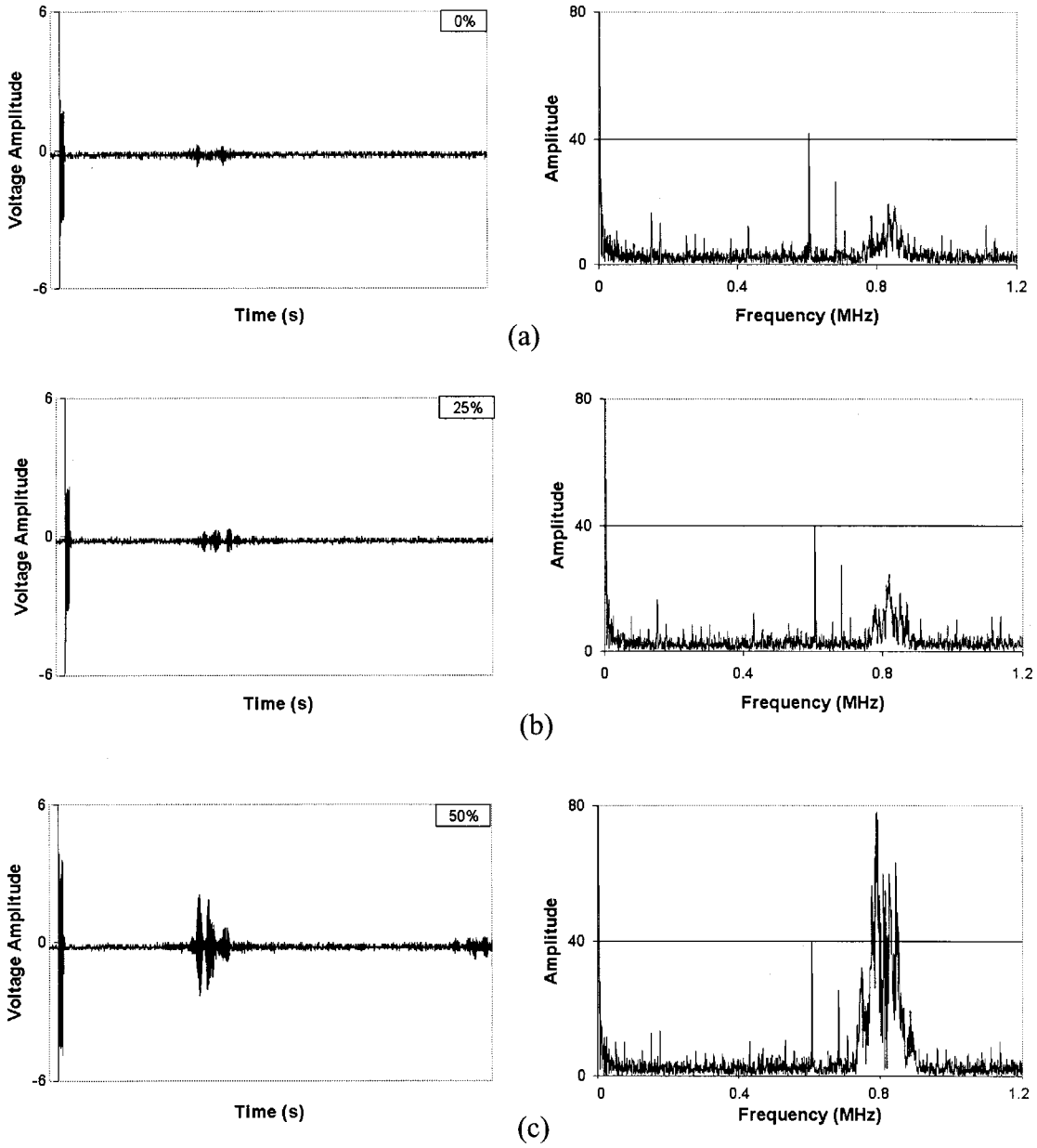


FIG. 9. $V(t)$ and $V(f)$ curves using the transmitter–receiver arrangement shown in Fig. 4(d): (a) 0%; (b) 25%; and (c) 50% delaminated specimens.

propagating a strong signal through the specimens with the transmitter–receiver arrangement shown in Fig. 4(a).

Figure 6 shows the $V(t)$ and $V(f)$ curves of the specimens using the transmitter–receiver arrangement shown in Fig. 4(a) with 24° incident angle. Three sets of curves correspond to $V(t)$ and $V(f)$ curves for 0%, 25%, and 50% delaminated specimens, respectively. These curves show that the percentage of delamination affects the received signal; hence, the delamination can be well quantified by this transmitter–receiver arrangement.

Figure 7 shows the $V(t)$ and $V(f)$ curves of the specimens using the transmitter–receiver arrangement shown in Fig. 4(b). Three sets of curves correspond to $V(t)$ and $V(f)$ curves for 0%, 25%, and 50% delaminated specimens, respectively. These curves show that the percentage of delamination can be quantified from this transmitter–receiver arrangement. These curves show a big jump in the signal amplitude as the delaminated region increases from 25% to 50%. The signals for 50% delamination are much stronger than the corresponding signals shown in Fig. 6.

Figure 8 shows the $V(t)$ and $V(f)$ curves of the specimens using the transmitter–receiver arrangement shown in Fig. 4(c). Three sets of curves correspond to $V(t)$ and $V(f)$ curves for 0%, 25%, and 50% delaminated specimens, respectively. These curves also show that the percentage of delamination can be well quantified from this transmitter–receiver arrangement. These signals are much stronger than those produced by other transmitter–receiver arrangements shown in Figs. 4(a), (b), and (d).

Figure 9 shows the $V(t)$ and $V(f)$ curves of the specimens using the transmitter–receiver arrangement shown in Fig. 4(d). Three sets of curves correspond to $V(t)$ and $V(f)$ curves for 0%, 25%, and 50% delaminated specimens, respectively. These curves show that only the high percentage of delamination ($>25\%$) can be detected by this transmitter–receiver arrangement. These signals are weaker than the sig-

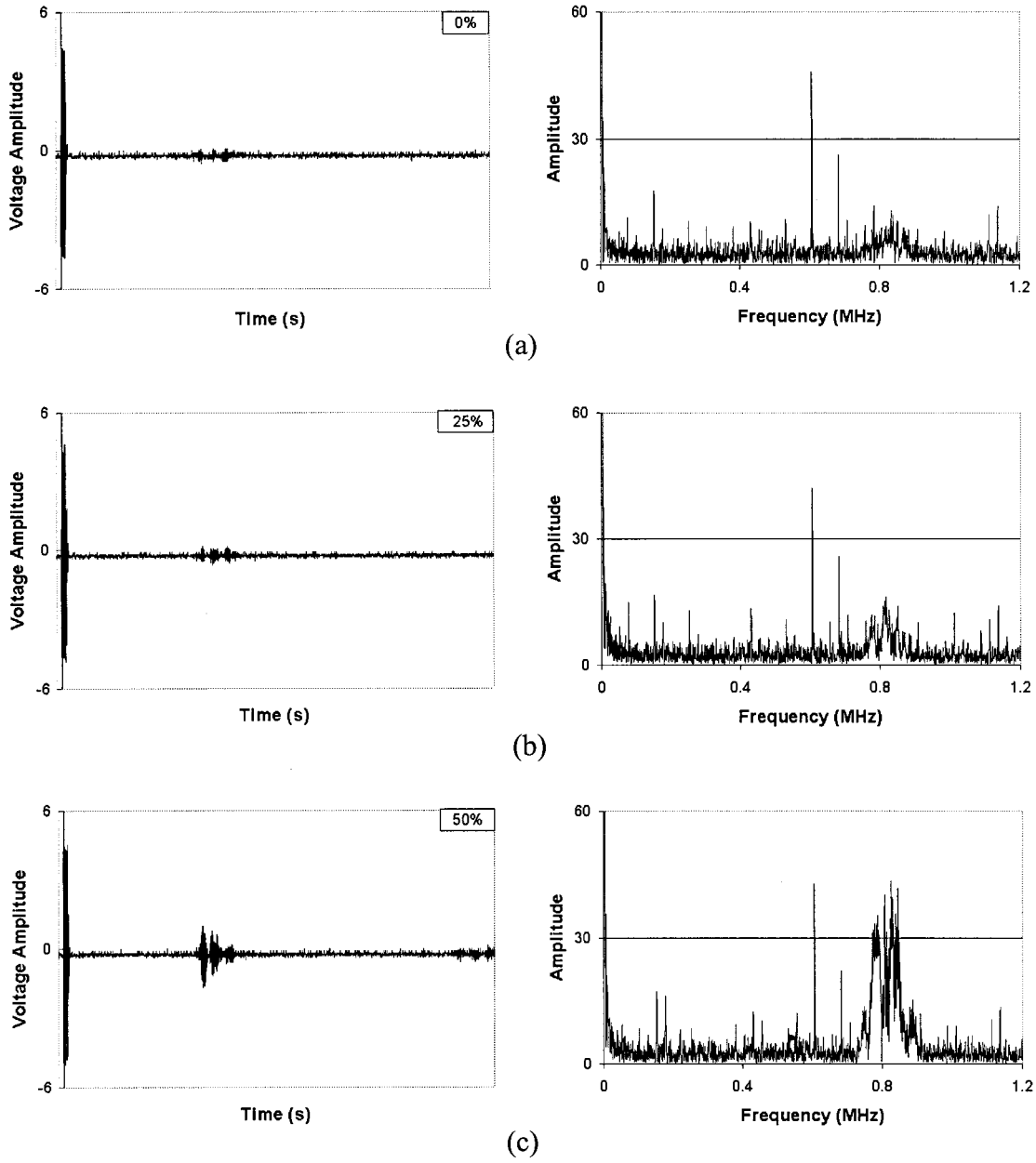


FIG. 10. $V(t)$ and $V(f)$ curves using the transmitter–receiver arrangement, shown in Fig. 4(d): (a) 0%; (b) 25%; and (c) 50% delaminated specimens. A paper is placed between the EMAT and steel bar.

nals obtained from the transmitter–receiver arrangements shown in Figs. 4(b) or (c).

Figure 10 shows the $V(t)$ and $V(f)$ curves of the specimens using the transmitter–receiver arrangement shown in Fig. 4(d). However, this time, a thin paper is placed between EMATs and a steel bar to investigate the effect of the inconsistency of the steel-bar surface. Three sets of curves correspond to $V(t)$ and $V(f)$ curves for 0%, 25%, and 50% delaminated specimens, respectively. These curves show that the high percentage ($>25\%$) of delamination can still be easily detected from this transmitter–receiver arrangement, although the $V(t)$ curves are weaker than those shown in Fig. 9. It indicates that small changes of the surface condition of a steel bar affect the strength of the $V(t)$ curves; however, in a relative scale similar conclusions can be drawn.

Figure 11 compares the $V(t)$ curves generated by the transmitter–receiver arrangements shown in Figs. 4(c) and (e). The left column shows the $V(t)$ curves when the arrangement of Fig. 4(c) is used and the right column corresponds to the $V(t)$ curves when the arrangement of Fig. 4(e) is used. Three rows correspond to the $V(t)$ curves for 0%, 25%, and 50% delaminated specimens, respectively. Curves in the left column show that the percentage of delamination can be well quantified from the transmitter–receiver arrangement shown in Fig. 4(c), as was the case in Fig. 8 for another experimental setup with the PZT–EMAT arrangement. Curves in the right column show that the difference in $V(t)$ curves generated by 0% and 25% delaminated specimens is not very clear. Hence, the sensor arrangement in Fig. 4(e), although it produces a stronger signal, is less effective than that in Fig. 4(c) for detecting smaller ($<25\%$) delamination.

To investigate the sensitivity of the experimental results

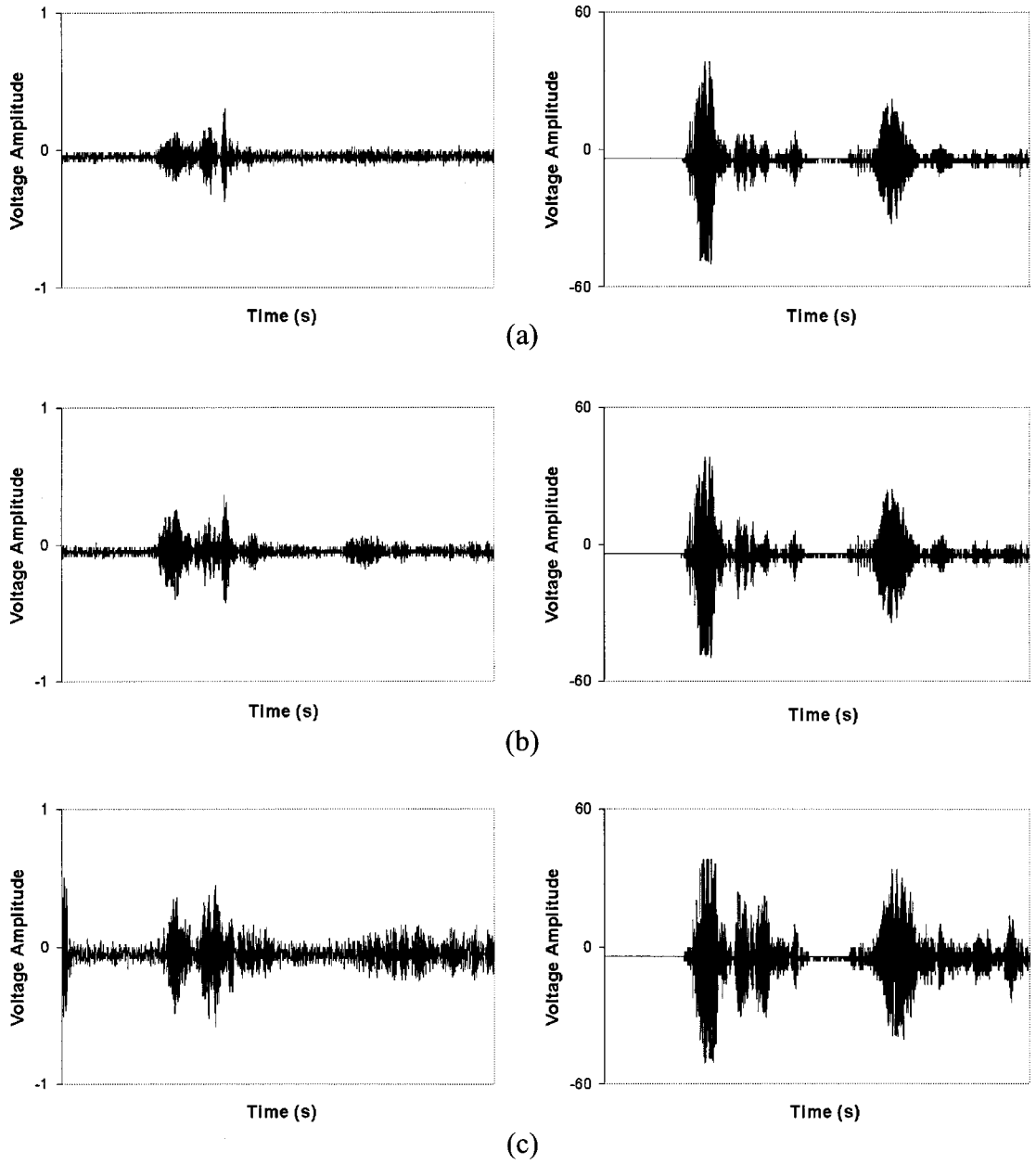


FIG. 11. $V(t)$ curves using the transmitter–receiver arrangements, shown in Fig. 4(c): (left column) and 4(e) (right column): (a) 0%; (b) 25%; and (c) 50% delaminated specimens.

to the steel-bar surface condition, the transmitter–receiver arrangements shown in Figs. 4(c) and (e) are tried out after inserting thin papers between the sensor and the steel bar. Figure 12 shows the $V(t)$ curves of the specimens using the transmitter–receiver arrangement shown in Fig. 4(c). However, this time, as stated above, a thin paper is placed between the EMAT and the steel bar. Two columns correspond to two different experimental results with two similar papers inserted. Figure 12 shows the consistency of the experimental results. Figure 13 shows the $V(t)$ curves generated by the transmitter–receiver arrangement shown in Fig. 4(e). Two columns correspond to two different experiments with similar thin papers placed between the receiver PZT and the steel bar. These curves clearly show inconsistency in the results. Hence, arrangement 4(c) is desirable for consistency in the experimental results.

IV. DISCUSSION

From the above experiments, the following conclusions can be drawn: (1) The strength of the received EMAT signals can be increased when PZT transducers are used as transmitters in direct contact with the steel bar as shown in Figs. 4(b) and (c). (2) Most transmitter–receiver arrangements shown here are good for detecting and quantifying the delaminated regions. (3) Small changes in the surface condition of the steel bar affect the amplitude strength of the $V(t)$ curves. (4) Surface finish can affect the consistency of the experimental results, but the inconsistency can be reduced by using EMAT as the receiver. (5) EMAT–EMAT and PZT–PZT combinations are good for detecting large ($>25\%$) delamination. However, for detecting small delaminations ($<25\%$) the PZT–EMAT combination is recommended.

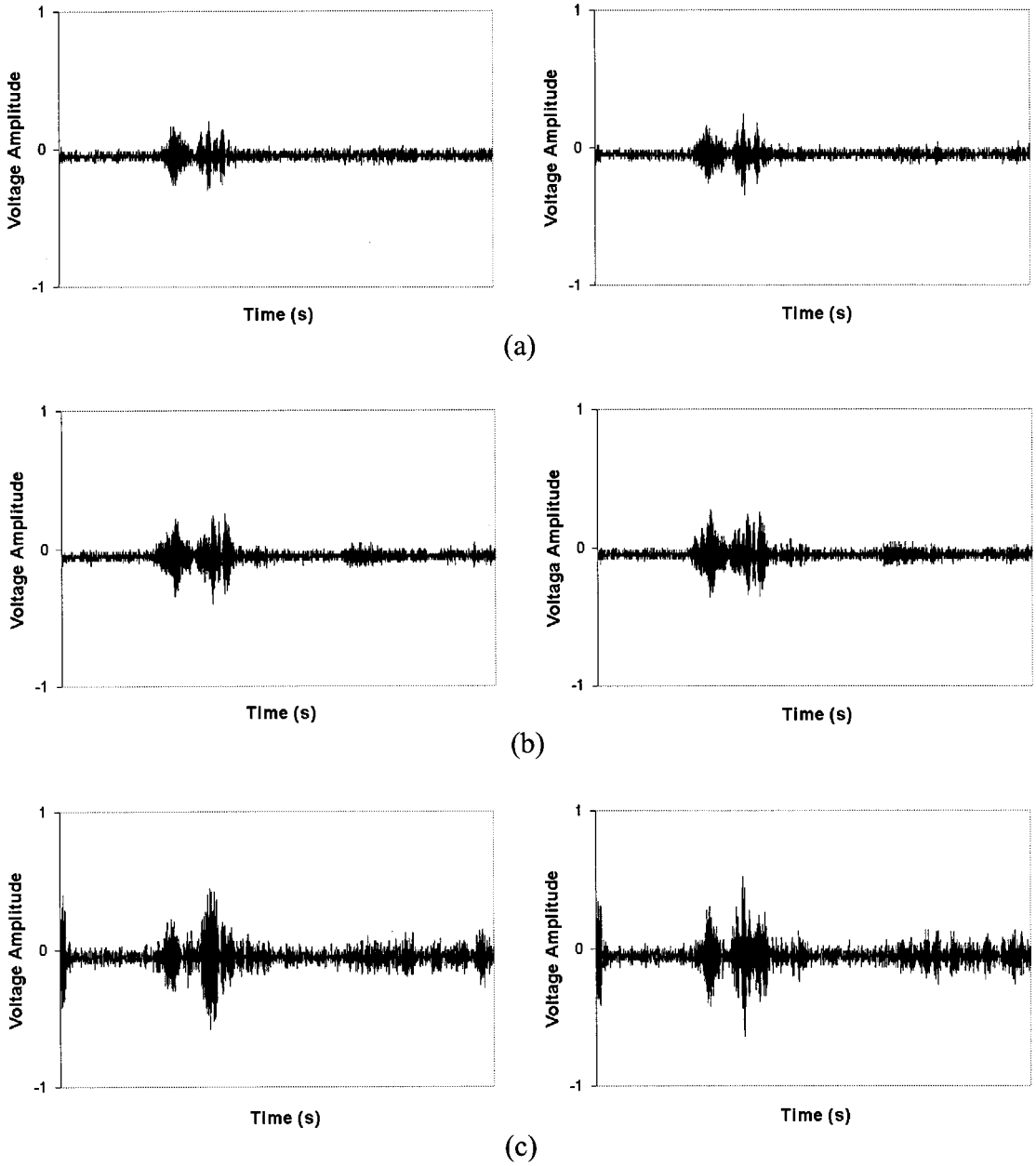


FIG. 12. $V(t)$ curves using the transmitter–receiver arrangement, shown in Fig. 4(c): (a) 0%; (b) 25%; and (c) 50% delaminated specimens. A paper is placed between the EMAT and steel bar. Two columns are for two different sets of experiments with two different but similar papers.

It should be noted here that this technique can generate a false alarm, if not properly used. This may happen because (1) the use of couplant under different environments can affect the experimental results, since the property of the couplant may change with the variation of the surrounding temperature and humidity; (2) surface roughness of the bar can change the clearance between the steel bar and EMAT; hence, it can change the magnitude of the $V(f)$ curves. In our investigation, we used papers between the receiving sensor and the steel bar to study the effect of the varying clearance; (3) rust can affect the surface roughness and bar dimension resulting in a change in the received signal amplitude.

Cracks in reinforced concrete can also affect the experimental results. Although there is no evidence of the existence of major cracks in the specimens that we studied, cracks at the concrete–steel interface should be properly modeled for accurately synthesizing the propagating guided waves. The interface cracks change the bonding condition; hence, less wave energy dissipates through the bonded surface. However, existence of minor cracks might not change the amplitude of the $V(f)$ curves significantly. This also induces us to decide what is the smallest percentage of detachment that can be detected by this technique. At this stage, it seems very difficult to answer this question since (1) a complete analytical study is yet to be carried out, and (2) sensitivity of the technique may be a function of the total length of concrete–steel bar interface inspected.

No data reduction operation has been applied to the $V(t)$ and $V(f)$ curves presented here. Appropriate signal processing can reduce the noise level in the received signal. This can be accomplished by placing a gate or a window on the $V(t)$

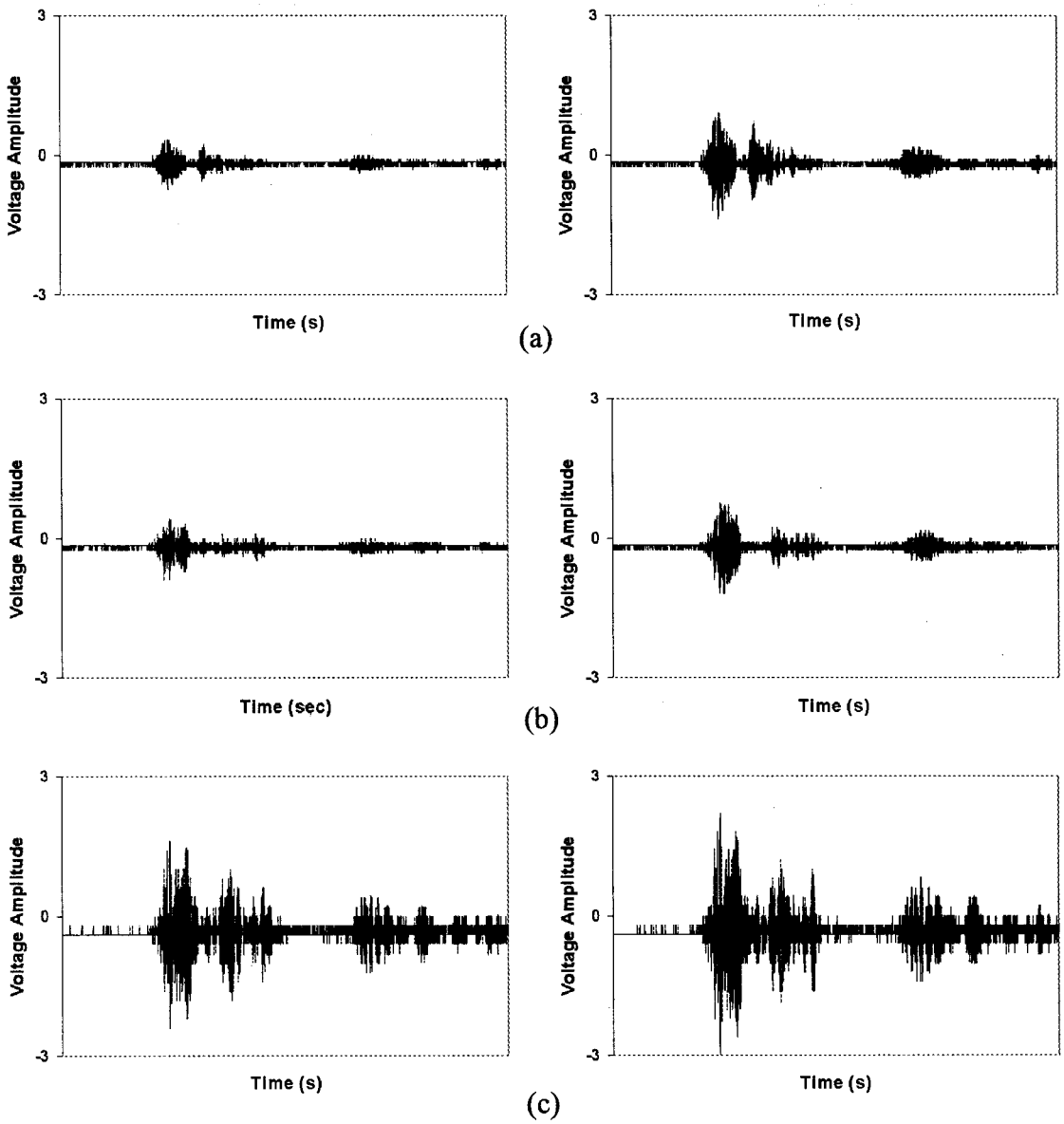


FIG. 13. $V(t)$ curves using the transmitter–receiver arrangement, shown in Fig. 4(e): (a) 0%; (b) 25%; and (c) 50% delaminated specimens. A paper is placed between the receiver PZT and the steel bar. Two columns are for two different sets of experiments with two different but similar papers.

curves instead of analyzing the entire signal. This gate control can help field operators to make right decisions. However, it may also complicate the operational steps for nonexpert field operators inspecting real structures.

Two limitations of this study should be mentioned here. First, deformed bars of different shapes are often used for reinforcement, while in this investigation only plain reinforcing bars have been considered. Second, most reinforcements are located inside the concrete; and it is difficult to apply this technique when reinforcing bars are not exposed. However, in many prestressed concrete structures, both ends of prestressed tendons are exposed; hence, this technique can be applied for locating interface delamination and voids in prestressed concrete.

V. CONCLUSION

An EMAT is introduced for inspecting the steel bar–concrete interface. To increase the strength of the received signal, PZT is used for generating the signal during the inspection. Various combinations of EMAT and PZT are tried out to find an efficient transmitter–receiver arrangement. $V(t)$ and $V(f)$ curves are obtained and their strengths are compared for different degrees of delamination. It is shown that several combinations are effective for steel bar–concrete interface inspection; however, PZT–EMAT combinations are most efficient for detecting small delamination (25% or less) at the steel bar–concrete interface.

ACKNOWLEDGMENTS

This research was financially supported by the NSF Grant No. CMS-9901221.

[1] T. C. Patton and D. K. Hsu, "Recent Developments of the Dripless Bubbler Ultrasonic Scanner," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thomson and D. E. Chimenti (Plenum, New York, 1998), Vol. 7, pp. 2045–2051.
[2] D. Guo and T. Kundu, "Special Sensors for Generating Lamb Waves in

Pipes,'' in *Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thomson and D. E. Chimenti (Plenum, New York, 1999), Vol. 18, pp. 1155–1162.
[3] D. Guo and T. Kundu, ''A New Sensor for Pipe Inspection by Lamb Waves,'' Mater. Eval. **58**, 991–994 (2000).
[4] D. Guo and T. Kundu, ''Lamb Wave Sensors for Detecting Wall Defects in Pipes,'' in *Nondestructive Evaluation of Highways, Utilities, and Pipelines IV* (Conference 3995B), edited by A. E. Aktan and S. R. Gosselin (Proceedings of SPIE NDE 2000, Newport Beach, CA, 2000), Vol. 3995, pp. 610–618.
[5] D. Guo and T. Kundu, ''A New Transducer Holder Mechanism for Pipe Inspection,'' J. Acoust. Soc. Am. **110**, 303–309 (2001).
[6] R. E. Green, ''Emerging Technologies for NDE of Aging Aircraft Structures,'' in Proceedings of the Workshop on Intelligent NDE Sciences for Aging and Futuristic Aircraft, edited by C. Ferregut, R. Osegueda, and A. Nunez, UTEP, El Paso, TX, pp. 267–278, 1997.
[7] H. Kwun and K. A. Bartels, ''Magnetostrictive Sensor Technology and Its Applications,'' Ultrasonics **36**, 171–178 (1998).
[8] L. Durris, L. Goujon, A. Pelourson, P. Gonnard, M. Brissaud, and C. Richard, ''Airborne Ultrasonic Transducer,'' Ultrasonics **34**, 153–158 (1996).
[9] C. B. Scruby and B. C. Moss, ''Non-Contact Ultrasonic Measurements on Steel at Elevated Temperatures,'' NDT&E International **26**, 177–188 (1993).
[10] H. Kwun and C. M. Teller, ''Detection of Fractured Wires in Steel Cables Using Magnetostrictive Sensors,'' Mater. Eval. **52**, 503–507 (1994).
[11] H. Ogi, T. Hamaguchi, and M. Hirao, ''*In-situ* Monitoring of Ultrasonic Attenuation During Rotating Bending Fatigue of Carbon Steel with Electromagnetic Acoustic Resonance,'' J. Alloys Compd. **310**, 436–439 (2000).
[12] D. A. Oursler and J. W. Wagner, ''Narrow-Band Hybrid Pulsed Laser/EMAT System for Non-contact Ultrasonic Inspection Using Angled Shear Waves,'' Mater. Eval. **53**, 593–598 (1995).
[13] M. Hirao, H. Ogi, and H. Yasui, ''Contactless Measurement of Bolt Axial Stress Using a Shear-Wave Electromagnetic Acoustic Transducer,'' NDT&E International **34**, 179–183 (2001).
[14] B. W. Maxfield, A. Kuramoto, and J. K. Hulbert, ''Evaluating EMAT Designs for Selected Applications,'' Mater. Eval. **45**, 1166–1183 (1987).
[15] K. Sawaragi, H. J. Salzburger, G. Hübschen, K. Enami, A. Kirihigashi, and N. Tachibana, ''Improvement of SH-Wave EMAT Phased Array Inspection by New Eight Segment Probes,'' Nucl. Eng. Des. **198**, 153–163 (2000).
[16] E. P. Papadakis, C. G. Oakley, A. Selfridge, and B. Maxfield, ''Fabrication and Characterization of Transducers,'' in *Ultrasonic Instruments and Device II*, edited by R. N. Thurston and A. D. Pierce (Academic, New York, 1999), pp. 43–134.
[17] B. Pavlakovic, M. Lowe, and P. Cawley, ''Guided Ultrasonic Waves for the Inspection of Post-Tensioned Bridges,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thomson and D. E. Chimenti (Plenum, New York, 1998), Vol. 17, pp. 1557–1564.
[18] Y. C. Jung, T. Kundu, and M. R. Ehsani, ''Internal Discontinuity Detection in Concrete by Lamb Waves,'' Mater. Eval. **59**, 418–423 (2001).
[19] W. B. Na, T. Kundu, and M. R. Ehsani, ''Lamb Waves for Detecting Delamination Between Steel Bars and Concrete,'' J. Comput. Aided Civil Infrastructure Eng. (in press).
[20] W. B. Na, T. Kundu, and M. R. Ehsani, ''Ultrasonic Guided Waves for Steel Bar Concrete Interface Testing,'' Mater. Eval. **60**, 437–444 (2002).

Factors affecting the performance of large-aperture microphone arrays

Harvey F. Silverman,[a] William R. Patterson III, and Joshua Sachar
LEMS, Division of Engineering, Box D, Brown University, Providence, Rhode Island 02912

(Received 11 June 2001; accepted for publication 1 March 2002)

Large arrays of microphones have been proposed and studied as a possible means of acquiring data in offices, conference rooms, and auditoria without requiring close-talking microphones. When such an array essentially surrounds all possible sources, it is said to have a large aperture. Large-aperture arrays have attractive properties of spatial resolution and signal-to-noise enhancement. This paper presents a careful comparison of theoretical and measured performance for an array of 256 microphones using simple delay-and-sum beamforming. This is the largest currently functional, all digital-signal-processing array that we know of. The array is wall-mounted in the moderately adverse environment of a general-purpose laboratory ($8\,\mathrm{m}\times 8\,\mathrm{m}\times 3\,\mathrm{m}$). The room has a T_{60} reverberation time of 550 ms. Reverberation effects in this room severely impact the array's performance. However, the width of the main lobe remains comparable to that of a simplified prediction. Broadband spatial resolution shows a single central peak with 10 dB gain about 0.4 m in diameter at the -3 dB level. Away from that peak, the response is approximately flat over most of the room. Optimal weighting for signal-to-noise enhancement degrades the spatial resolution minimally. Experimentally, we verify that signal-to-noise gain is less than proportional to the square root of the number of microphones probably due to the partial correlation of the noise between channels, to variation of signal intensity with polar angle about the source, and to imperfect correlation of the signal over the array caused by reverberations. We show measurements of the relative importance of each effect in our environment.
[DOI: 10.1121/1.1471895]

PACS numbers: 43.58.Vb, 43.38.Kb, 43.58.Kr [SLE]

I. INTRODUCTION

There has been modest, but steady, research into microphone-array technology over the last ten years or so. Several special workshops have been held.[1–4] Much of the work has been either theoretical[5,6] or based on simulations using mathematical models for the array's performance and the characteristics of the environment.[7,8] In most experimental papers, results have been based upon relatively small arrays having from two to twenty-one microphones.[9–13]

Large arrays of microphones have been proposed and studied[7,14–16] as a possible means of acquiring data in offices, conference rooms, and auditoria without requiring close-talking microphones. When such an array essentially surrounds all possible sources, it is said to have a large aperture. Large-aperture arrays have attractive properties of spatial resolution and signal-to-noise enhancement.[7] This paper presents a careful comparison of theoretical and measured performance of an array of 256 microphones using simple delay-and-sum beamforming. This is the largest currently functional, all digital-signal-processing array of which we are aware.

This work uses delay-and-sum beamforming for several reasons. Our ultimate goal is to build a high-quality system for speech recognition and recording that can track multiple moving talkers in the presence of noise. Any such system must first combine microphone outputs to maximize the signal from the desired source relative to competing noise or other sources. Of the known methods for doing this, delay-and-sum is the simplest. The necessary filter can be constructed only knowing the effective source position, which can be derived strictly from the acoustical data. Precisely because it does not rely on the details of the acoustic environment, delay-and-sum is also robust, showing consistent, reproducible SNR gain and source isolation. Any algorithm of this kind tends to introduce artifacts as parameter values change to accommodate a moving talker. For delay-and-sum, it is straightforward to predict exactly when such effects will appear and to compensate for them. Finally, more powerful methods, such as matched filtering,[17–19] approximate inverse filtering,[20] and null steering[21] have a history of only working well in tightly constrained situations. In looking closely at the factors that affect how well delay-and-sum actually works, we hope to explore ways to overcome the limitations of other algorithms.

The array is wall-mounted in the moderately adverse environment of a general-purpose laboratory ($8\,\mathrm{m}\times 8\,\mathrm{m}\times 3\,\mathrm{m}$). The room has a T_{60} reverberation time of 550 ms, and there are from 8 to 44 discrete reflections above -20 dB in each channel. We present performance results for a large real-time system that can support up to 512 microphones. The hardware, called the Huge Microphone Array (HMA), has been fully described in Refs. 22 and 15 and the software and algorithms in Refs. 16, 14, and 23. The front-end hardware has 32 modules, each of which does analog-to-digital conversion (16 bit) for 16 microphones and multiplexes that

[a]Electronic mail: hfs@lems.brown.edu

data onto a fiber optic link to a central processing console. Data are transmitted in short bursts every 25.6 ms. The link is bidirectional with the uplink used to maintain both sample and frame synchronization. The console has 96 floating point DSPs that can give a sustained total computation rate of approximately 4 GFlops. All processors have access to all the microphone data and can exchange data with any other processor. Data transfer is directed by a PC workstation and the system output is available as either PC files or as outputs from a set of 12 DACs.

After some mathematical models are developed, we present computed results for the current arrangement of microphones but for an idealized environment in which (1) there are no reverberations or background noise, (2) microphones are point receivers at the nominal positions in the real array, and (3) the source location is known exactly.

The remainder of the paper explains and brings together data measured from the physical system with this simulation. One major problem in the comparison is the absence of any reverberant energy in the computed results even though the physical array is not in an anechoic chamber. A method to isolate the direct wave was proposed by Heyser,[24] in which a narrowband filter is used with a chirp. Called time-energy-frequency analysis, it has been described in detail in Refs. 25–29. However, the technique used in this paper does not require the use of a narrowband filter and takes advantage of our entirely digital system to ensure very precise synchronization of repetitive test signals. Short chirps allow one to separate direct from indirect paths while long bursts of noise serve as a basis for measurements of room transfer functions. High signal-to-noise ratio measurements are possible by exploiting the precise repetition to reduce uncorrelated noise. We believe that this work explains the sources of performance variation for large-aperture array systems in real-world environments very well.

II. MATHEMATICAL MODELS

The acoustic sources for the present data were a series of small loudspeakers that could be moved accurately about the room. However, the main use of the array is for human speech so the source model must be general enough for both speakers and people. For both source types, the measured signal at point $\mathbf{r}_m$ depends on the position of the source $\mathbf{r}_s$, the orientation of the source, and its excitation. In turn, both the position and orientation may be functions of time, so we write the signal as $s(\mathbf{r}_m-\mathbf{r}_s(t),\hat{r}_o(t),\hat{r}_e(t),t)$.

Here we have specified orientation in terms of two unit vectors attached to the source, one, $\hat{r}_e$, extending forward from the center of the source (whether from the cone or mouth as the case may be) and the other, $\hat{r}_o$, marking the nominally vertical axis. We simplified our measurements by always having speakers and talkers stand upright and by leaving the speaker in one position for the duration of a particular measurement. Commonly, people are observed to remain stationary for times long compared to the room reverberation time (in our case of $T_{60}=550$ ms). With these assumptions, the signal in the time domain becomes $s(\mathbf{r}_m-\mathbf{r}_s,\hat{r}_e,t)$, while its equivalent Fourier transform is $S(\mathbf{r}_m-\mathbf{r}_s,\hat{r}_e,\omega)$. In the absence of scattering and reverberation and at distances several times the lateral extent of the source, S takes the form of spherical waves with phase and amplitude modified by a function of the relative direction between the emission direction vector and the source orientation vectors. Suppose we measure the signal directly in front of the source at unit distance and compensate for the wave propagation from the nominal source to the measuring point. This forms a practical working definition of the array excitation $S'(\omega)$ such that

$$S(\mathbf{r}_m-\mathbf{r}_s,\hat{r}_e,\omega)=\frac{S'(\omega)H^S(\theta,\phi,\omega)\exp\left(\frac{-j\omega|\mathbf{r}_m-\mathbf{r}_s|}{v}\right)}{|\mathbf{r}_m-\mathbf{r}_s|}, \tag{1}$$

where v is the velocity of sound. The *radiation pattern function*, $H^S(\theta,\phi,\omega)$, is complex and describes the modification of the source spectrum with emission angle. Its arguments, θ and ϕ, are the usual spherical angles describing the direction of emission, $\mathbf{r}_m-\mathbf{r}_s$, relative to the axes defined by the source orientation vectors $\hat{r}_o$ and $\hat{r}_e$. The values $\theta=90°$ and $\phi=0°$ describe propagation directly forward from the source along $\hat{r}_e$. There is a subtle assumption in writing H^S that the angular spectral distribution is time independent. This is clearly appropriate for a loudspeaker, but people vary the size of the mouth opening during speech. They also have appreciable emission from the nose as well as the mouth, and the relative importance of each changes with time. Meuse has recently shown[30] by both experimental measurements and by modeling with spherically baffled ideal sources that mouth-related effects are very difficult to detect in a measured $H^S(\theta,\phi,\omega)$. Here we consider them small second-order effects that we neglect.

The signal received by any particular microphone, indexed by n, on the array can now be calculated by considering all possible paths from the source. Let $H_n^R(\theta,\phi,\mathbf{r}_s,\mathbf{r}_n,\omega)$ be the propagation transfer function for a signal that ultimately reaches microphone n at $\mathbf{r}_n$ after starting in the direction defined by θ and ϕ. This function encapsulates the effects of frequency-dependent scattering but not the decrease in amplitude with distance imposed by the inverse square law, as that is included in the explicit distance factors. The microphone signal is

$$M_n(\mathbf{r}_s,\mathbf{r}_n,\omega) = S'(\omega)\oint\oint \sin(\theta)\,d\theta\,d\phi \times\frac{H_n^R(\theta,\phi,\mathbf{r}_s,\mathbf{r}_n,\omega)H^S(\theta,\phi,\psi)\exp\left(\frac{-j\omega|\mathbf{r}_n-\mathbf{r}_s|}{v}\right)}{d_{ns}(\theta,\phi)}, \tag{2}$$

where $d_{ns}(\theta,\phi)$ is the path length along the particular path associated with a given emission direction. Note that the double integral is what is usually referred to as the room transfer function (RTF), but here it is explicitly a function of the source orientation through the emission angle dependencies of H^S and H^R. In situations with clear lines-of-sight from the source to all microphones, it is useful to separate

the direct-path contribution to the microphone signal from the integral and to express the result as

$$M_n = S'(\omega)\left[\frac{H^S(\theta_{ns},\phi_{ns},\omega)\exp\left(\frac{-j\omega|\mathbf{r}_n-\mathbf{r}_s|}{v}\right)}{|\mathbf{r}_n-\mathbf{r}_s|} + \oint\oint_{\text{indirect paths}} \sin(\theta)\,d\theta\,d\phi \times \frac{H_n^R(\theta,\phi,\mathbf{r}_s,\mathbf{r}_n,\omega)H^S(\theta,\phi,\omega)\exp\left(\frac{-j\omega d_{ns}(\theta,\phi)}{v}\right)}{d_{ns}(\theta,\phi)}\right],$$

where θ_{ns} and ϕ_{ns} are the pointing angles from source to microphone n relative to the source orientation axes.

Because of the geometry of the environment, H_n^R will be impulsive with respect to the angles θ and ϕ, that is only a discrete, though large set of values of (θ_i,ϕ_i) will contribute to the integral. The double integral becomes a discrete sum in which some terms represent identifiable early discrete reflections. The remaining terms, even though still a discrete sum, produce the effect of later general reverberant energy that is more like noise than information. Recognizing the discrete nature of the terms making up the integral, we write,

$$M_n = S'(\omega)\times\left[\frac{H^S(\theta_{ns},\phi_{ns},\omega)\exp\left(\frac{-j\omega|\mathbf{r}_n-\mathbf{r}_s|}{v}\right)}{|\mathbf{r}_n-\mathbf{r}_s|} + \sum_i R_{ns}^i\right], \tag{3}$$

where

$$R_{ns}^i \equiv \frac{H_n^R(\theta_i,\phi_i,\mathbf{r}_s,\mathbf{r}_n,\omega)H^S(\theta_i,\phi_i,\omega)\exp\left(\frac{-j\omega d_{ns}(\theta_i,\phi_i)}{v}\right)}{d_{ns}(\theta_i,\phi_i)} \tag{4}$$

is the transfer function for the ith signal path in this channel. In practice, it is possible to determine $H^S(\theta_{ns},\phi_{ns},\omega)$ and $\Sigma_i R_{ns}^i(\omega)$, but not H_n^R or the angles (θ_i,ϕ_i). The radiation pattern function $H^S(\theta,\phi,\omega)$ is an important limit to how well signals from widely separated microphones can add coherently. Section VI includes measurements of this function for one of our speakers, while Sec. VII discusses measurements of room reverberation in terms of the intensity and delay associated with some of the discrete reflections of the room and of the total reverberant energy that does not contribute to the addition of signals in the beamformer.

A delay-and-sum beamformer first multiplies the Fourier transform of the output of each microphone by a complex delay factor equivalent to the distance between a given aiming point and the microphone itself. That is, if $\mathbf{r}_A$ is the aiming point, let $\tau_n^A \equiv |\mathbf{r}_n-\mathbf{r}_A|/v$ be the propagation delay from the aiming point to microphone n, let $\tau_n^S \equiv (|\mathbf{r}_n-\mathbf{r}_S|)/v$ be propagation delay from the source to microphone n and multiply m_n by $e^{j\omega\tau_n^A}$. Then a per-channel weighting coefficient, W_n, is applied and the resulting signals are added together to form the array output. In addition to the desired signal, each channel receives some noise, N_n, from competing acoustic sources and from electrical effects in the microphones and analog-to-digital converters. Thus the output of a N microphone array is given by

$$O_{\text{array}} = \sum_{n=1}^{N} W_n e^{j\omega\tau_n^A}\left[S'(\omega)\left(\frac{H^S(\theta_{ns},\phi_{ns},\omega)e^{-j\omega\tau_n^S}}{|\mathbf{r}_n-\mathbf{r}_s|} + \sum_i R_{ns}^i(\omega)\right) + N_n\right]. \tag{5}$$

III. THE ARRAY AND TESTING ENVIRONMENT

Our current array of 256 microphones, which has been used for all simulations and measurements, is in an 8.4 m $\times$ 8.4 m normal laboratory room in which the floor is hard tile and the ceiling is cement 3 m above the floor. The ceiling has approximately rectangular boxlike cavities that are about 4 m $\times$ 1 m $\times$ 0.3 m in a regular pattern. As shown top-view of Fig. 1, the room is subdivided, but the smaller rooms have walls only about 2 m high so the top is open. The loss from scattering into the smaller rooms over the walls and from the large foam surfaces of the arrays themselves make the measured T_{60} reverberation time consistent with approximate calculations based on Sabine's law (Ref. 31, p. 718). The array has eight wall-mounted panels each 1.34 m $\times$ 0.67 m with 32 omnidirectional microphones placed on a randomly selected subset of the nodes of a 30 mm grid, which assures a minimum separation between microphones. The microphones are countersunk into the 40-mm-thick foam panel. Note that the x, y, z coordinate system has its origin in the center of the first panel.

IV. COMPUTED RESULTS FOR AN IDEAL SYSTEM

In an earlier paper, we studied the optimization of microphone placement for a large array using a simplified ideal model, comparing three general positioning strategies.[7] This ideal model ignores reverberation and noise and assumes an isotropic point source. To obtain satisfactory agreement with measured results, we have found it necessary to retain at least the angular variation of the radiation pattern function. As the microphones are at roughly the same height as the source, there is no appreciable dependence on θ_{ns}. With these assumptions, the array output, Eq. (5), simplifies to

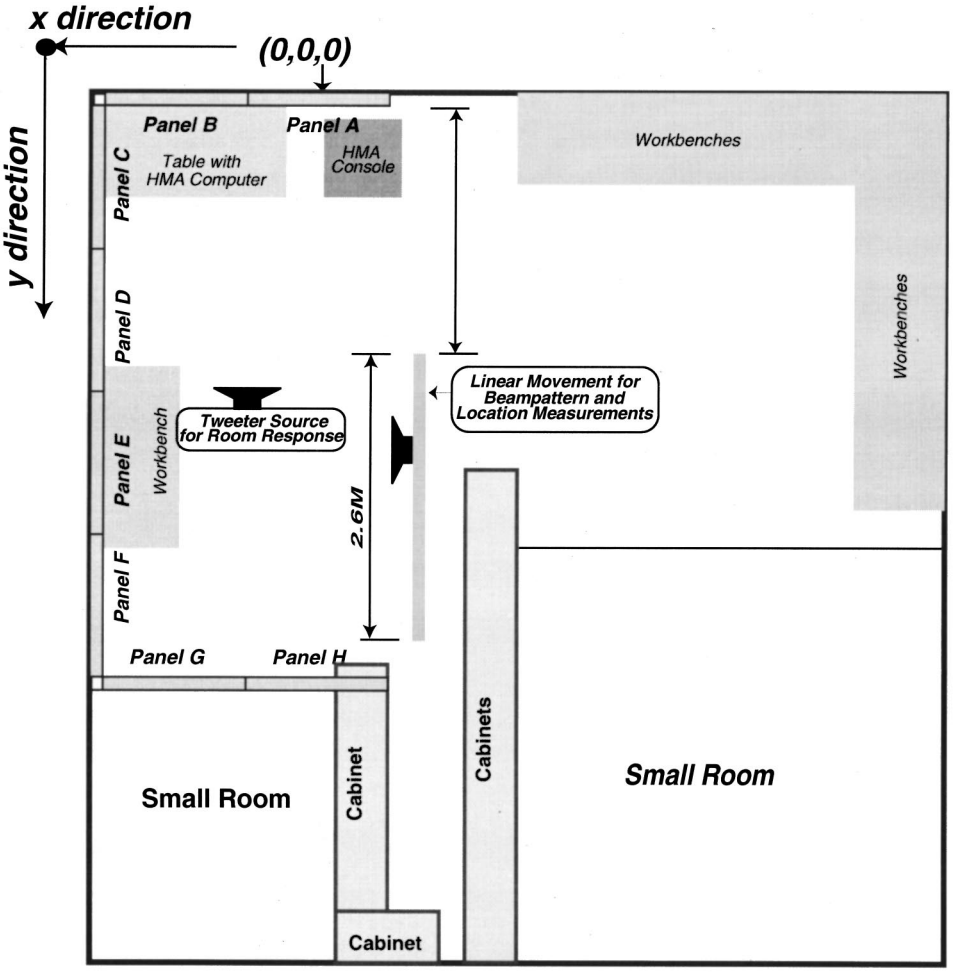


FIG. 1. Plan view of the laboratory showing the array and the adjoining spaces—Z direction is from floor to ceiling.

$$O_{\text{array}} = S'(\omega) \sum_{n=1}^{N} W_n \frac{H^S(\phi_{ns}, \omega) e^{j\omega(\tau_n^s - \tau_n^A)}}{|\mathbf{r}_n - \mathbf{r}_s|}. \tag{6}$$

The sum in Eq. (6) represents the complex *beam pattern*[32] function $B(\mathbf{r}_s, \mathbf{r}_A, \omega)$ at a sinusoidal driving frequency ω, i.e.,

$$B(\mathbf{r}_s, \mathbf{r}_A, \omega) \equiv \sum_{n=1}^{N} W_n \frac{H^S(\phi_{ns}, \omega) e^{j\omega}(\tau_n^s - \tau_n^A)}{|\mathbf{r}_n - \mathbf{r}_s|}. \tag{7}$$

The beam pattern may be thought of as the measured response of an array system that has been aimed to a single location $\mathbf{r}_A$ as the source itself is moved through the room. [Alternatively, one could also consider the sum in Eq. (6) as the *steered response* function in which the array is steered to various locations with the source fixed. These are equivalent in the ideal case, but are not the same in a nonideal environment due to different reverberation patterns.] We typically express the magnitude of B in dB,

$$\beta(\mathbf{r}_s, \mathbf{r}_A, \omega) \equiv 10 \log |B(\mathbf{r}_s, \mathbf{r}_A, \omega)|^2. \tag{8}$$

Equation (8) may be used to generate ideal, steady-state sinusoidal responses for the array. In Fig. 2, we show the magnitude response at three locations, 20, 50, and 100 cm away from a fixed aiming point for an isotropic source $[H^S(\phi_{ns}, \omega) = 1]$. These curves are representative of most similar calculations, showing a series of decreasing sidelobes separated by pronounced nulls. Null separation decreases with distance away from the aiming point and the sidelobe behavior is qualitatively similar to $(\sin u/u)^2$ where $u = \omega x/v$. At 20 cm displacement of the source from the aiming point, frequencies below 500 Hz are not affected very much by the beamformer, while frequencies above about 1300 Hz are all attenuated more than 12 dB. At 50 cm, only frequencies below 200 Hz are unaffected by the beamformer, and frequencies over 500 Hz have at least 12 dB of attenuation. At 100 cm, the frequency increment (100 Hz) of the plot is probably too coarse to show all details. Nevertheless, there is about 15 dB minimum attenuation of all frequencies over 500 Hz.

All signals of real interest are intrinsically broadband and any useful performance metric must take this into account. One way to compute such a metric is to replace $B(\mathbf{r}_s, \mathbf{r}_A, \omega)$ in Eq. (8) with a weighted sum over a range of frequencies. This is equivalent to computing the mean power for a hypothetical test signal with a known power spectral density normalized to give 0 dB response for a source at the aiming point. The variation of this quantity with position is the effective spatial resolution of the array. Let $F(\omega_i)$ be the normalized spectrum of the test signal. Then the array response is

$$\beta_f(\mathbf{r}_s, \mathbf{r}_A) = 10 \cdot \log \left[\sum_{i=1}^{L} F(\omega_i) \cdot |B(\mathbf{r}_s, \mathbf{r}_A, \omega_i)|^2 \right]. \tag{9}$$

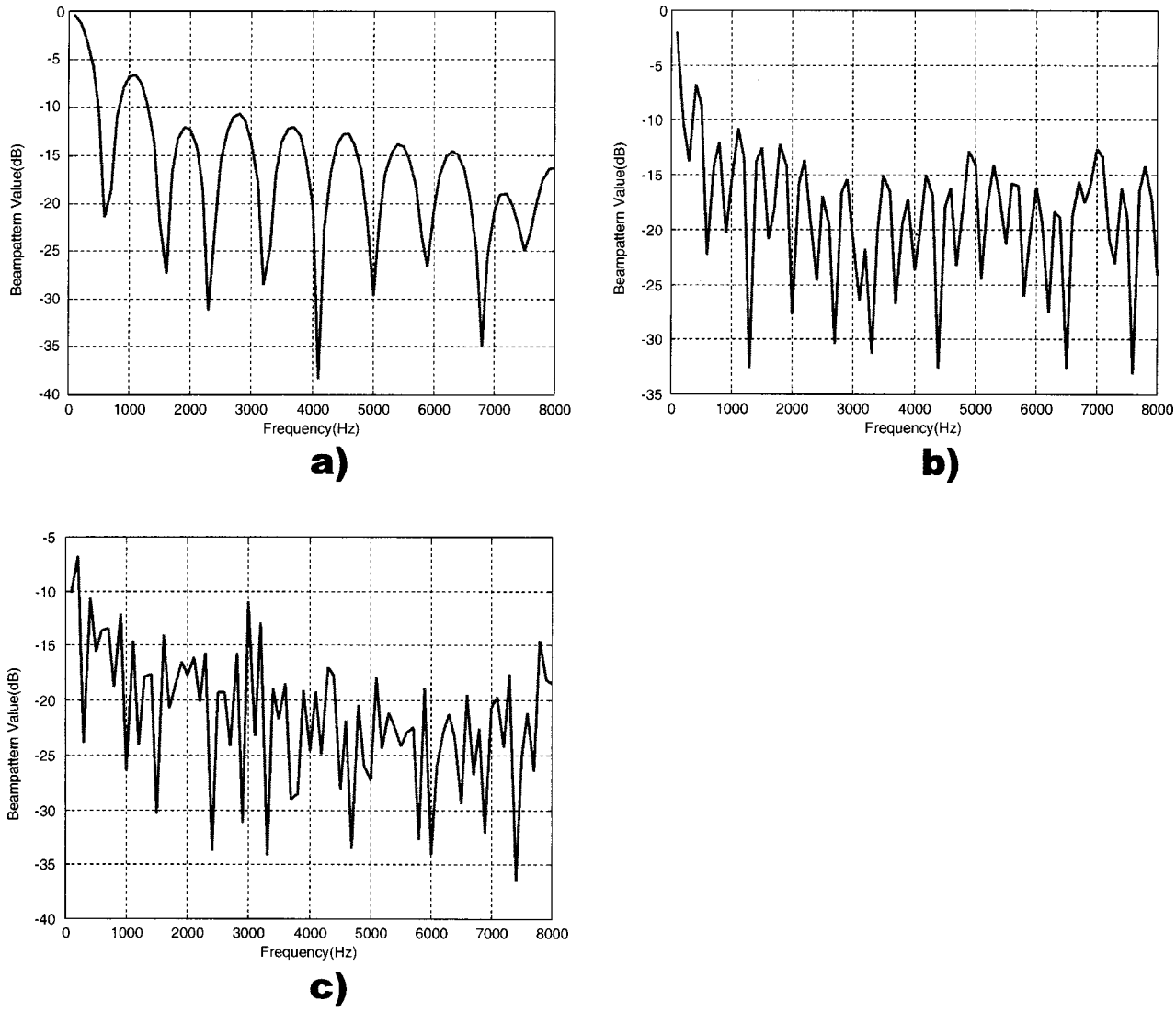


FIG. 2. Magnitude spectrum of the beamformer at a point on the Line of Fig. 1 that is (a) 20 cm, (b) 50 cm, and (c) 100 cm from the aiming point.

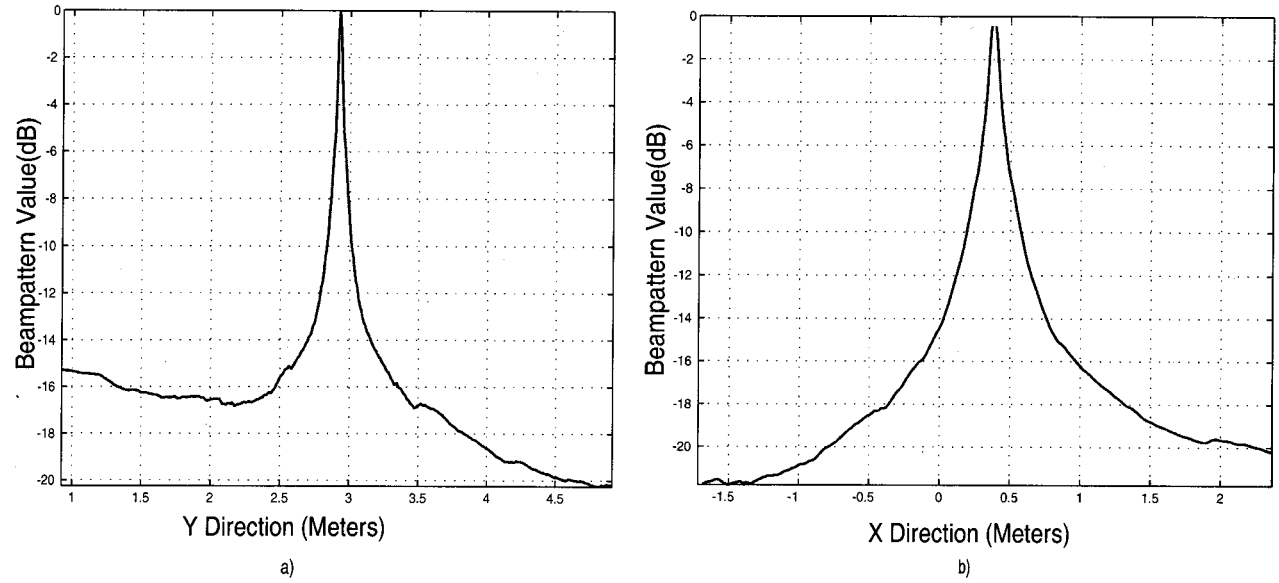


FIG. 3. Computed ideal response along (a) line parallel to the *Y* axis through the source and (b) line parallel to the *X* axis through the source using Eq. (9) with uniform weighting over the frequency range 500–6000 Hz.

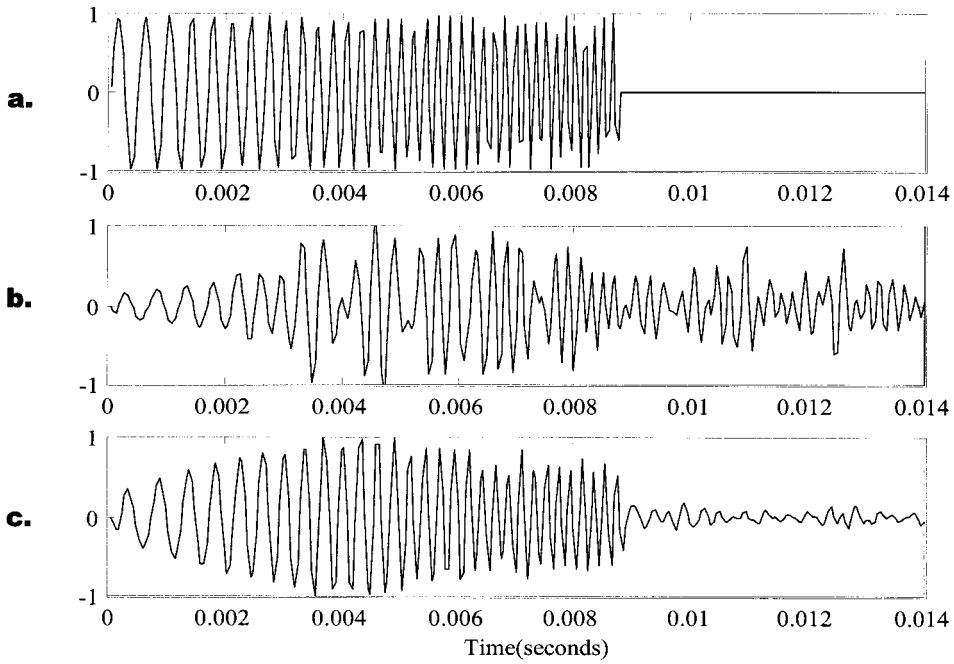


FIG. 4. (a) Original chirp, (b) received signal at a single microphone, after averaging over 10 000 repetitions, (c) beamformed sum of the 256 microphone signals.

In Figs. 3(a) and (b), the overall beamformer performance is shown over lines parallel to the y and x axes, respectively, through the source location for $F(\omega_l)$ chosen as a rectangular function giving unity weighting over the frequency range from 500 to 6000 Hz. The data demonstrate the effects of averaging over frequency and the broadband performance that might be expected. The null and peak behavior for sinusoidal sources is gone, yet an overall beam pattern is clear. The single main lobe width is quite narrow, measuring about 25 cm at -12 dB in Y and 65 cm at -12 dB in X and ultimate attenuation is better than 15 dB. As one would expect, for the x direction in which the aperture is about half that of y, the -12 dB spatial bandwidth is about twice that of Y.

V. MEASURING REVERBERATION CHARACTERISTICS WITH A SYNCHRONIZED TEST SIGNAL SYSTEM

Our hardware system includes a synchronized test signal system (STSS) that can repetitively produce an arbitrary programmable wave form synchronized to the received data with an accuracy guaranteed to about 50 ns, three orders of magnitude smaller than the 50 μs sampling time. This implies that time data from many pulses can be successfully averaged. Thus, enhanced signal-to-noise measurements can be made that effectively remove all uncorrelated electrical and acoustic noise.

One such signal is the 8.9 ms chirp swept linearly in frequency from 2 to 6.5 kHz shown in Fig. 4(a) as it is applied to the transducer. A typical microphone response is shown in Fig. 4(b), while the beamformed signal from all 256 microphones, uniformly weighted, appears in (c). Ten thousand chirp repetitions (this number of repetitions is only advantageous in highly stable environments, and many fewer would be used in most practical situations) were used with a delay of 1.2 s between them to guarantee that any residual power from one chirp was essentially zero before the next chirp. This averaging technique reduced any effects from background noise by at least 40 dB, ultimately allowing nearly 80 dB of observable reverberation effects. Note the following.

(1) Following the end of the source chirp, reverberant energy remains very strong, only slightly weaker than from the excitation. It is readily detectable until about 440 ms.

(2) At the beginning of the response, only the direct path signal is received, i.e., no reflections from this chirp have yet reached the microphone. The beginning of the response is an amplitude-modulated swept sinusoid for about 3 ms. The source transfer function determines the shape of the amplitude envelope.

(3) After about 3 ms, a first reflection is detected and the shape of the output no longer resembles the original chirp.

The advantage of the uniformly weighted beamformer is clear from Fig. 4(c). Reverberation effects after the end of the chirp signal have been reduced by more than 12 dB. The chirp itself is simply the electrical stimulus as amplitude modulated by the source transfer function. The positive effect of the weighted-delay-and-sum beamforming, even when uncorrelated noise is essentially removed, is clear; the reverberant energy in the desired signal has been reduced significantly.

VI. THE SOURCE RADIATION PATTERN FUNCTION

A 3 in., textile-dome, mid-range speaker, a Vifa model D75MX-30-08 (Ref. 33) was used for most of the experiments in this paper. [This speaker has six holes (0.9 cm diameter) drilled into the magnet area, allowing pressure relief and a significant back wave.] It was suspended on slender rods to minimize local reflection patterns. Measurements were taken by placing it on a tripod with a calibrated rotating head 65 cm from an omnidirectional microphone in the same room but as far from the walls and ceiling as possible. Acoustical foam was placed on portions of the floor and

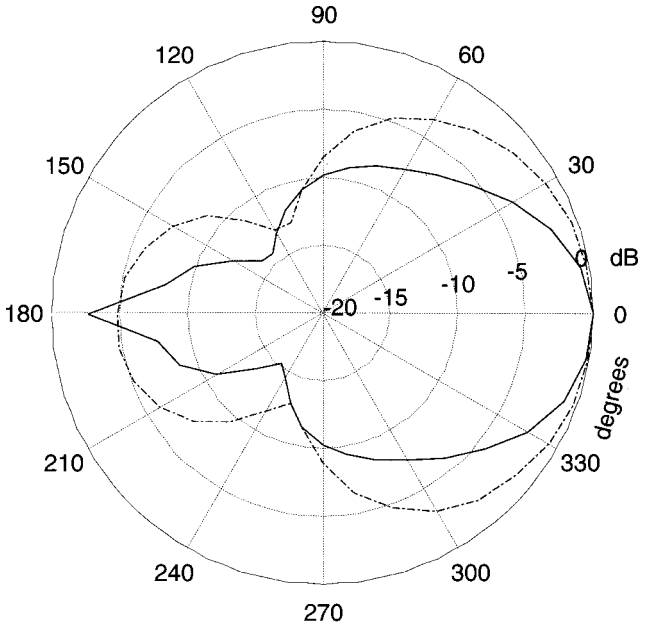


FIG. 5. Energy in the first 4 ms of excitation for a D75MX-30-08 midrange speaker as a function of azimuth angle—solid line chirp with mean frequency 3000 Hz, dashed line chirp with mean frequency 550 Hz.

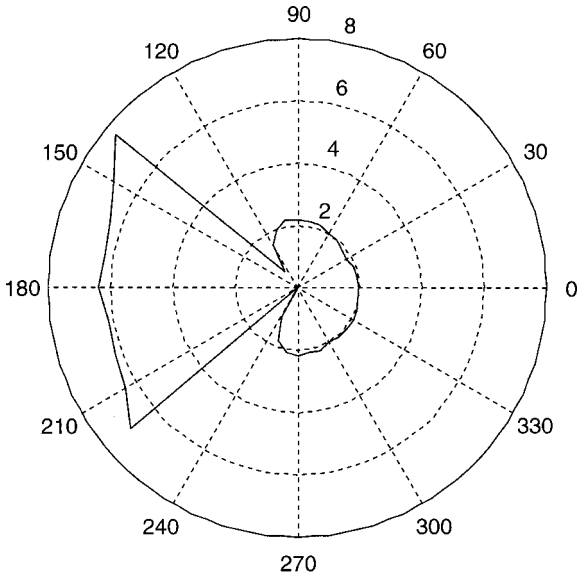


FIG. 6. Group delay variation in centimeters for beginning of linear 8.9 ms chirp 2000–6000 Hz from D75MX-30-08 midrange speaker, i.e., delay variation for mean frequency 2200 Hz.

ceiling to attenuate the most important early reflections. The STSS was used as the source, averaging over 100 chirps to reduce noise by 20 dB. In Fig. 5, the power in the direct wave is plotted for two different chirps. The solid line is for the first 4 ms (2000–4000 Hz) of an 8.9 ms linear chirp and the dashed line is for the first 4 ms (500–600 Hz) of a 72 ms logarithmically swept chirp.

The patterns are quite directional and frequency dependent. The solid line, representing an average frequency of 3000 Hz, shows attenuation greater than 6 dB for off-axis angles greater than 45° and reaches a loss of greater than 15 dB at the minimum response angle of 130°. The dashed curve for 600 Hz has broader lobes but still shows 6 dB attenuation beyond 80°. These curves are a direct measurement of the magnitude of the source radiation pattern function $H^S(\theta,\phi,\omega)$. Each is an average over frequency of $|H^S(\theta,\phi,\omega)|^2$ for $\theta=90°$, as ϕ rotates over 360°.

All our algorithms assume point sources and calculate delays based on the position of that point relative to the microphones. For macroscopic sources, it is possible that deviations in time of arrival as a function of angle might invalidate this assumption and degrade performance. To test the hypothesis, we used the data from the 8.9 ms chirp (2–6.5 kHz) and measured the time of arrival of the first zero crossing within the pulse. This is equivalent to measuring the group delay of the system around 2.2 kHz. We did a best fit of this data to a simple cosine function to remove the effect of the rotation axis not coinciding with the effective emission point. Figure 6 shows the delay variation with angle in terms of an equivalent distance. Over the front 220°, the total variation is less than 0.4 cm, which is within measurement error. Thus, in that range of angles the assumption of a point source appears to be quite justified.

The back of this particular source is vented through a set of six small holes, which give rise to an inverted back wave. Near 120°–130° off axis, the front and back waves partially cancel, causing the minima observed in the amplitude patterns of Fig. 5. That cancellation also introduces appreciable delay variations. In the approximately ±30° from directly behind the speaker, the inverted signal again appears to emanate from a point, but that point is about 4 cm behind the effective source position of the front wave. Obviously if one wished to include signals from behind this source in forming a beam, compensation would be necessary both for the inversion and for the appreciably different (4 cm or 0.12 ms) delay.

VII. THE ROOM TRANSFER FUNCTION

Our room is not acoustically treated. Since reverberant energy has serious implications for applications of this array, particularly for speech recognition, we sought to characterize the room effects carefully. Using the 8.9 ms chirp of Fig.

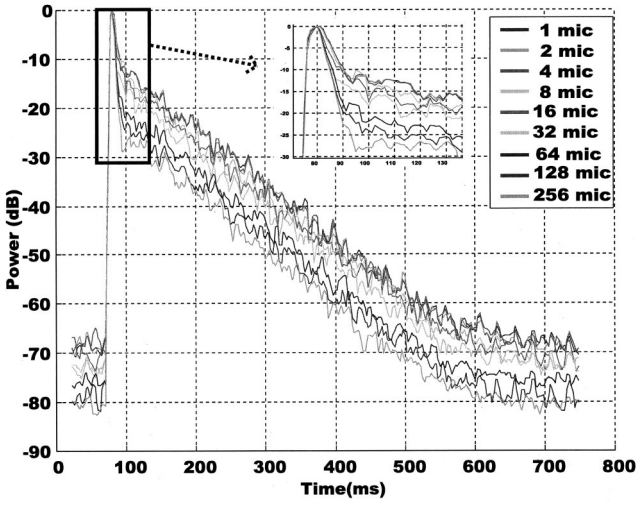


FIG. 7. Power response for various numbers of microphones for large-aperture array.

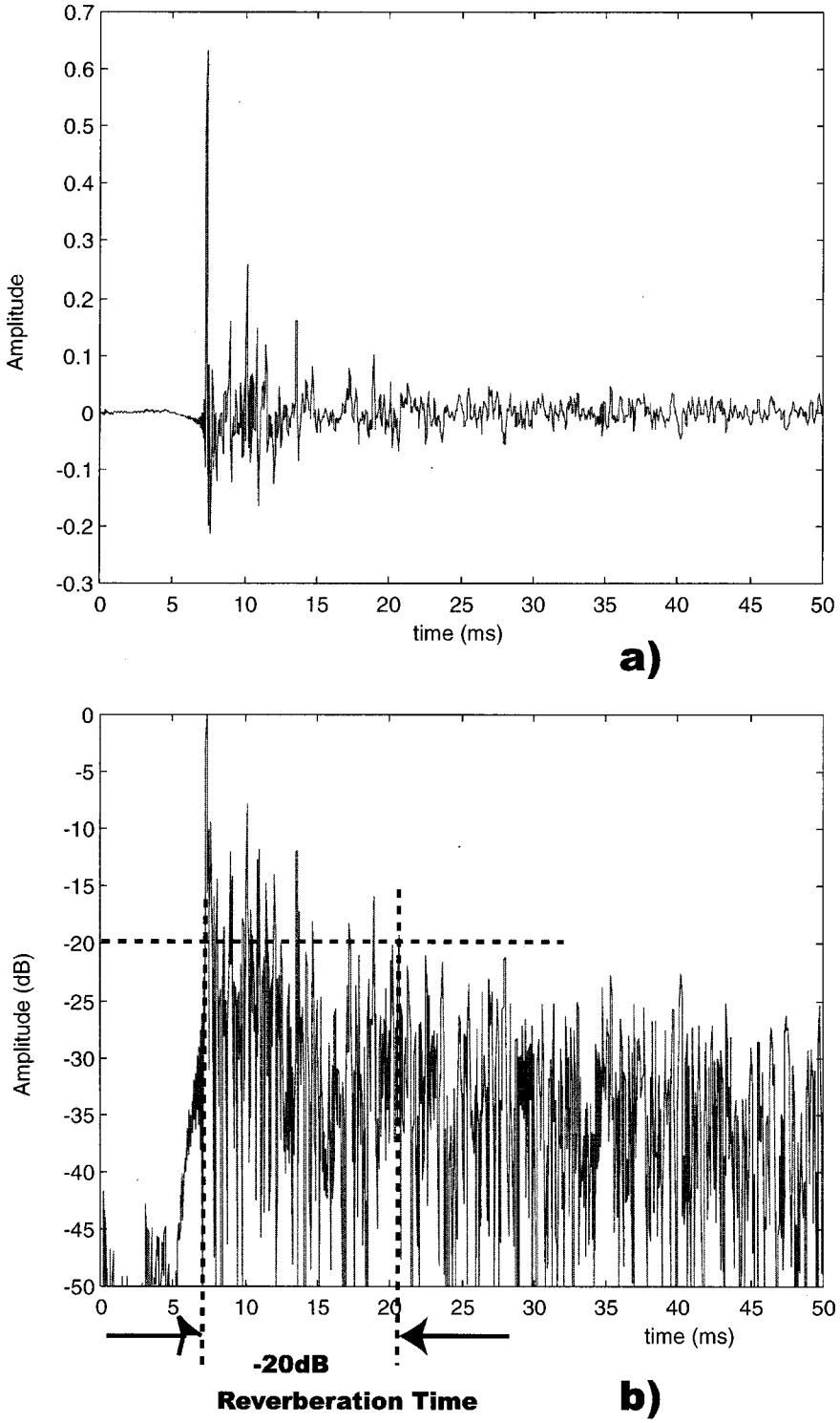


FIG. 8. First 50 ms of room impulse response computed from measurements for a single microphone directly in front of the source; (a) amplitude, (b) magnitude in decibels.

4(a), we measured reverberation decay for 0.78 s after the chirp for single microphones and for beamformed sums of 2, 4,8,16,...,256 microphone sets. Averaging data for 1000 chirps separated by 1.2 s for each data set allowed us to eliminate the effect of random noise and to reach almost 80 dB dynamic range. In Fig. 7 the relative powers (in dB) for 108-point (5.4 ms) segments are plotted with an advance of 50 points (2.5 ms). The data are normalized such that the peak of each experiment is 0 dB and occurs at about 80 ms after the onset of excitation. (Much of that 80 ms is processing latency.) The inset shows more detailed responses near the peaks. The reverberation time of a room, T_{60}, is defined[31] as "the time for the sound pressure level to decrease by 60 dB after the source has been cut off." Conventionally, this measurement is done using the slope of a log-linear plot. Our technique allows us to make this measurement in a noisy environment and even for the lower reverberation energy characteristic of a beamformed signal. There are a number of observations that may be made using Fig. 7 and its higher resolution inset graph of the first 50 ms of these data.

(1) The slope of the highest of the curves (one micro-

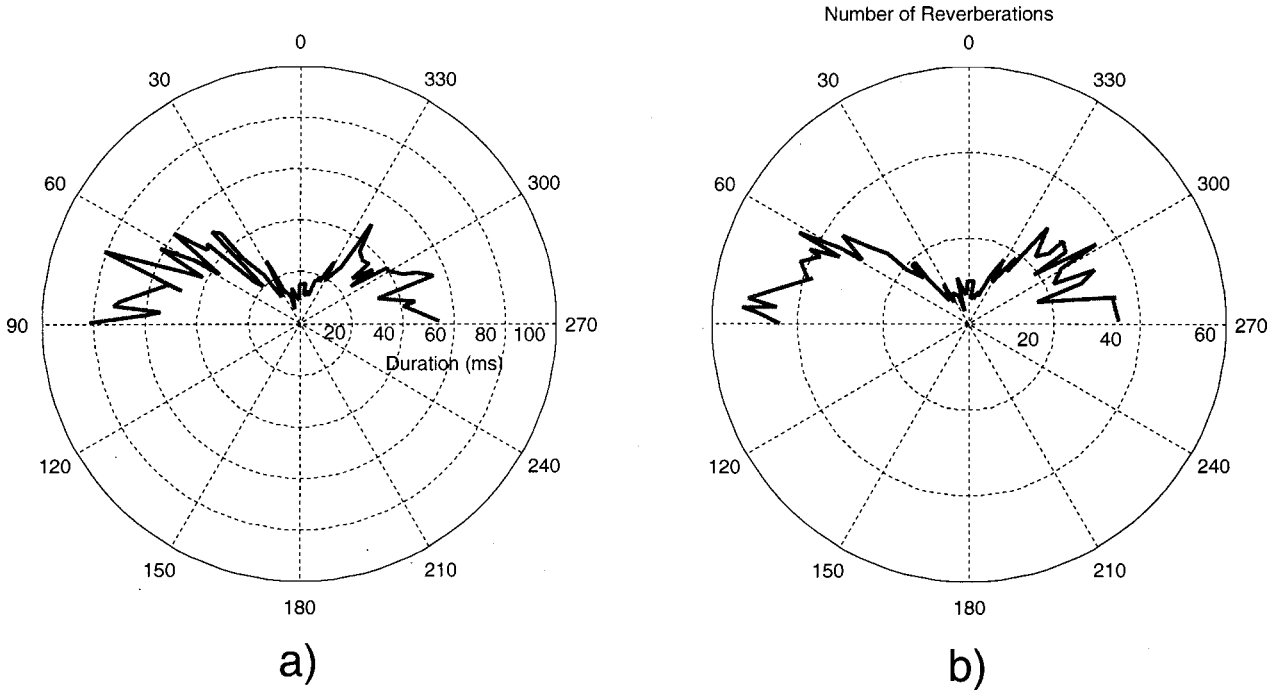


FIG. 9. Polar plots for impulse response using a sample of 62 microphones around the array; (a) number of reflections above −20 dB, (b) over −20 dB reverberation time.

phone) indicates that the room reverberation time is about 550 ms. This curve is essentially the same for any of the omnidirectional microphones making up the array.

(2) The inset figure quantifies the results from Fig. 4(c). For a single microphone, the response decays slowly after the chirp has been shut off. For 256 microphones, one can observe the expected spread due to the window used to compute power, but the residual signal 10 ms after the chirp is attenuated by 25 dB, versus about 12 dB for the single microphone.

(3) Beamforming does not alter T_{60}, but attenuates the entire reverberant energy by about 15 dB. This shortens the time for the reverberant energy to fall below any given threshold by approximately 100 ms.

A second experiment was conducted to obtain an accurate representation of the room acoustics. It used the STSS to generate 10 000 repetitions of 2.3 s of white noise, band limited to about half the sampling frequency (10 kHz). A reference microphone of the same type as used in the array was mounted 23 cm in front of the front plane of the source transducer. An effective room impulse response was calculated for each microphone by dividing the DFT of its response by the DFT of the referent signal and taking the inverse DFT of the ratio. This procedure removes the transducer transfer function and the microphone frequency response but leaves the effects from the source radiation function, room acoustics, and low-level reverberation in the referent channel. DFTs with 64 k points (3.27 s) were used so that the entire signal including all measurable reverberant energy is captured. Figures 8, 9, and 11 show typical results.

In Fig. 8, the first 50 ms of the impulse response for a microphone about 2.7 m directly in front of the transducer is shown. In Fig. 8(a) the impulse response is shown on a linear scale, and in Fig. 8(b) the magnitude is plotted in dB with the direct wave impulse set to 0 dB. About twenty discrete reflections over −20 dB may be seen over the first 13 ms following the direct wave. Looking at the 7.5 ms prior to the direct wave—the time it takes for the wave front to travel from the reference microphone to the array microphone—it is clear that the STSS technique has effectively eliminated the uncorrelated noise.

Figure 9 summarizes individual microphone impulse-response characteristics for a subset of 62 microphones spaced uniformly about the array. A −20 dB reverberation

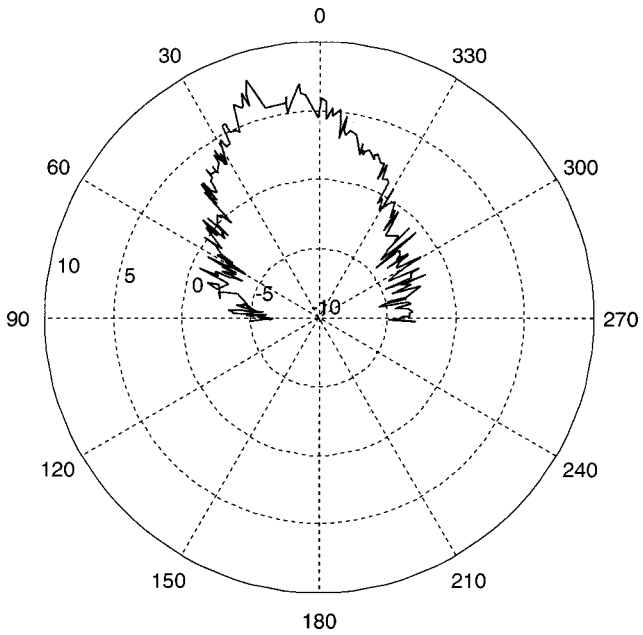


FIG. 10. Polar plot of the ratio of direct to reverberant signal energies.

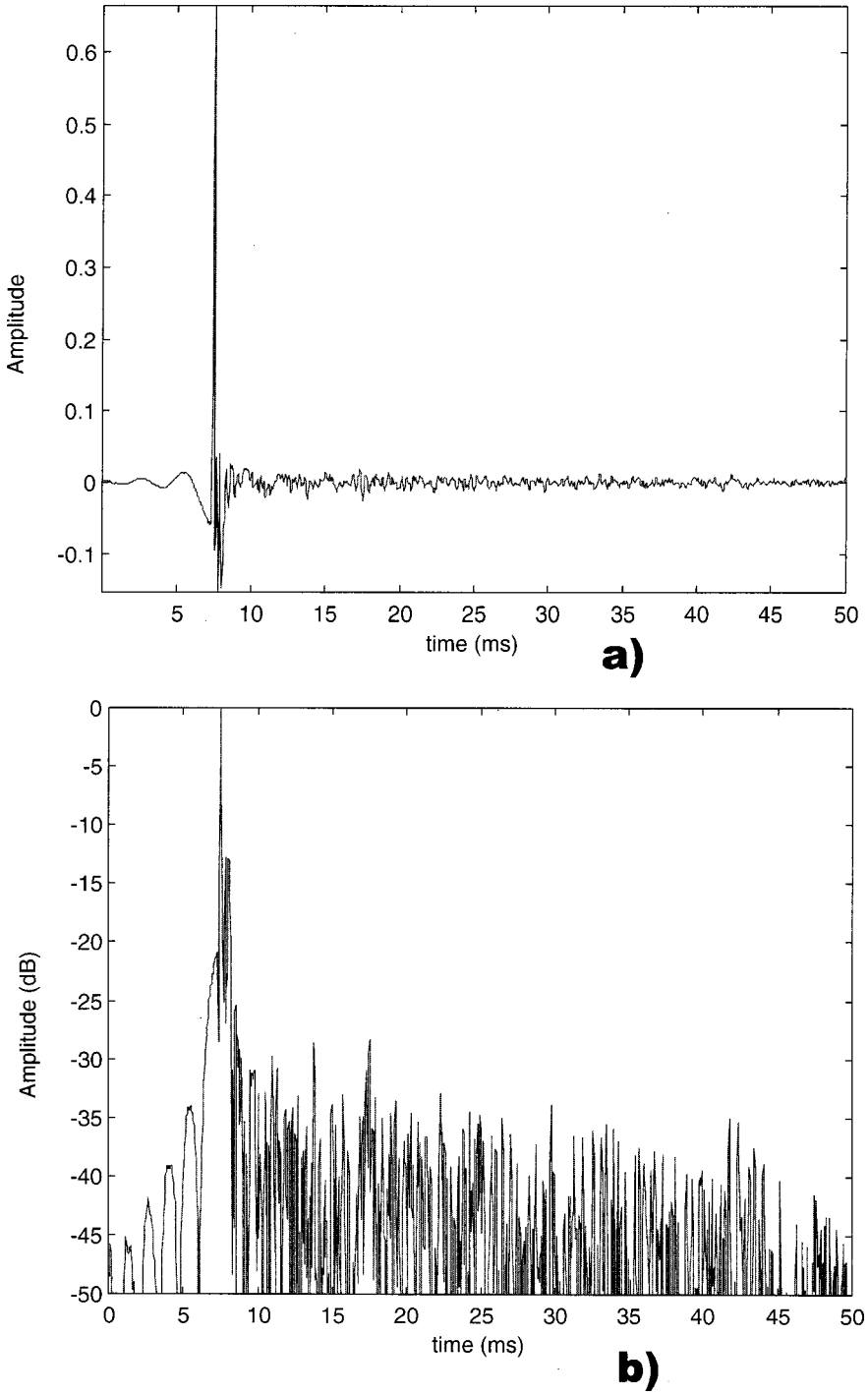


FIG. 11. First 50 ms of room impulse response computed from measurements for the beamformed signal; (a) amplitude, (b) magnitude in decibels.

time is defined pictorially in Fig. 8(b) as the time of the last reflection peak greater than 20 dB below the direct peak. In Fig. 9(a), this interval of major discrete reflection peaks is plotted as a function of ϕ, the angle between the emission direction to each microphone and the normal to the front of the transducer. The apparent reverberation time increases by a factor of 5 or so when the microphones are at angles of $\pm 90°$. Figure 9(b) is a similar polar plot counting the number of discrete reflections above -20 dB. It is interesting to see that there are fewer than ten reflections for microphones directly in front of the source, while four or five times that number for microphones near $\pm 90°$ away from the source normal. The reason for both results is that the radiation pattern function makes the direct signal 12 dB weaker at a $\pm 90°$ orientation. The room reflections, however, are approximately isotropic, that is, the reverberant energy and number of peaks per channel is about constant around the array. This is apparent in Fig. 10, which shows the ratio of direct signal energy to reverberant energy and which essentially reproduces the shape of the radiation pattern function shown in Fig. 5.

Figure 11 shows the impulse response for the first 50 ms of the beamformed signal. All the significant reflections appear substantially reduced until all reverberation peaks have

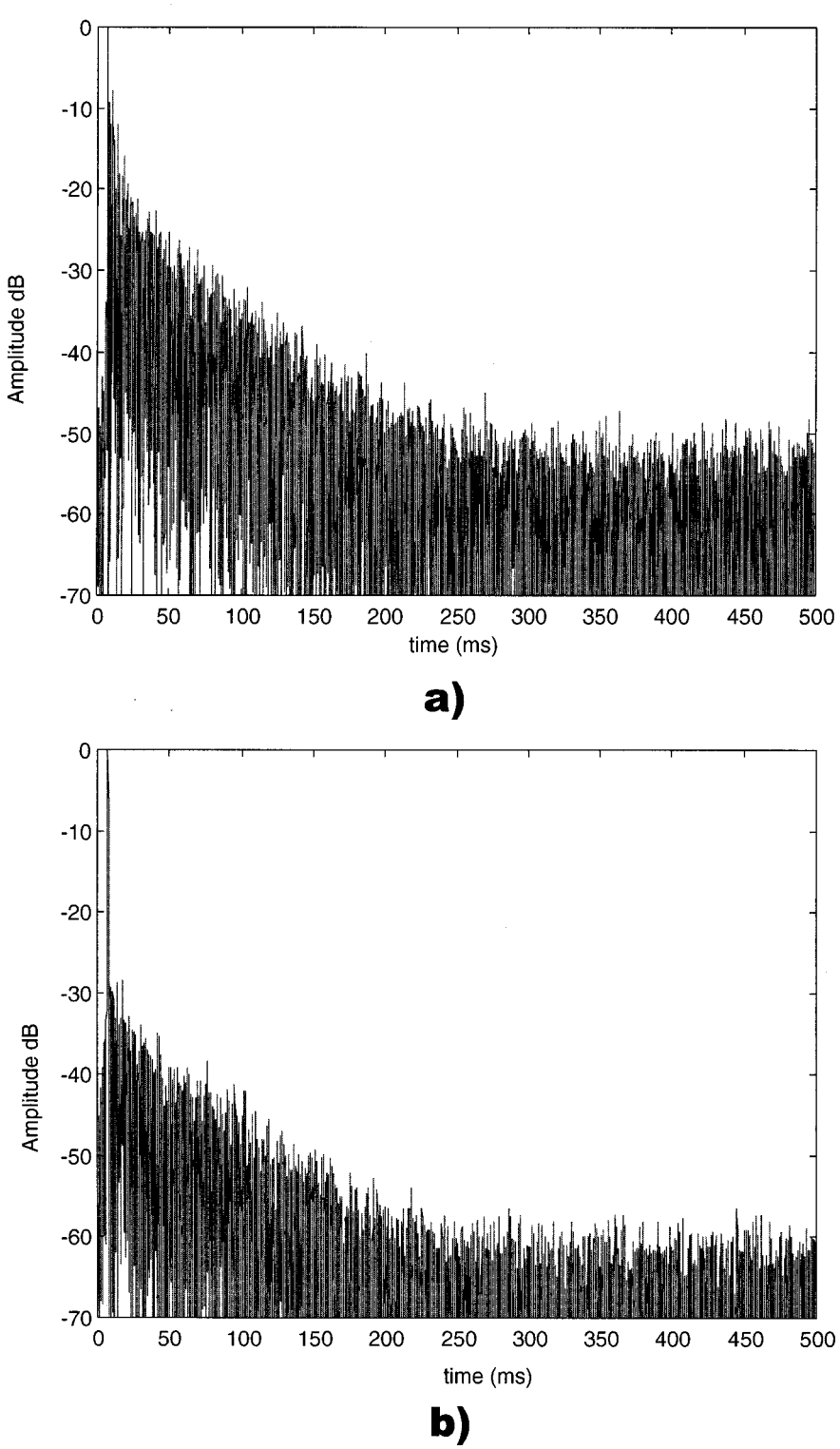


FIG. 12. First 500 ms of room impulse response computed from measurements; (a) magnitude (in dB) for a single microphone, (b) magnitude (in dB) for the beamformed signal.

been reduced to no higher than −28 dB relative to the direct wave peak value. While two peaks remain that are very close to the time of the direct wave, they are too near the direct wave in time to be bounces off walls, floor, etc. Rather, we believe that these are due to reflection off the walls on which the panels are mounted. Although the microphones are mounted on 6 cm of foam, some energy still reaches the wall and is reflected back, a 20 cm (0.5 ms) round trip.

The magnitudes in dB of the single microphone and the beamformed impulse response are shown in Figs. 12(a) and (b), respectively. The STSS has allowed us to eliminate most of the background noise for the single microphone, but the noise floor is about −50 dB judging from the data near the end of the plot. Here again, the reverberant energy decays exponentially, consistent with $T_{60} \approx 550$ ms. The improvement from the beamformer is quite evident in Fig. 12(b). The direct wave impulse is clearly 28 dB above any residual reverberation and remaining reverberation energy becomes unmeasurable beyond 250 ms. The noise floor relative to the peak of the direct wave is now about −60 dB.

VIII. COMPARISON OF MEASURED PERFORMANCE TO MATHEMATICALLY SIMULATED PERFORMANCE

Ultimately, there are two closely related purposes that a microphone array serves. The first is to distinguish one source from another by providing spatial resolution beyond that of a single microphone, and the second is to improve the signal-to-noise ratio (SNR) for a given source relative to both background noise and reverberation effects. In applying an array to speech recognition, it is also important that the frequency response be reasonably flat so spectral features are retained. Also, the SNR relative to reverberations must be high so speech data are more like the product of a close-talking microphone. Using the STSS as a signal source, we have studied how the spatial resolution of our array is affected by reverberation and by either of two choices of weighting coefficients. We then looked at the SNR achieved at the aiming point for the same two sets of weights relative to a single, optimal reference microphone. Again the STSS source allowed us to separate the ratio of the direct-signal-energy-to-reverberation-energy from the ratio to background-noise energy. Finally, the same data give some hint of the direct-wave frequency response.

Two data sets were taken for spatial resolution measurements by aiming the array at a fixed location, (0.379, 2.92, 0.007 m), and by then moving the source along a line parallel to the *Y* axis for 2.6 m centered about that aiming point. Previous work[16,14] showed that measurement along a single line in the room is adequate to characterize resolution. The difference between the two data sets was the choice of weighting coefficients; (1) equal weights or (2) weights that were assumed to optimize the SNR at the aiming point. The signal was the same 8.9 ms chirp (linearly swept from 2 to 6.5 kHz) used earlier. The source moved in increments of 1 cm near the aiming point and 10 cm away from it. At each location, the output of the beamformer was averaged for 1000 chirps, attenuating background noise by -30 dB and essentially eliminating it. The system recorded 0.8 s of this averaged data. In Sec. VIII A we examine this data on different time scales to separate the effects of reverberation, and we compare those measurements with the theory of an ideal noiseless and reverberation-free system from Sec. IV. In Sec. VIII B, we attempt a similar analysis for the SNR and frequency response at the aiming point.

A. Spatial resolution

For each recording, we made three calculations; (1) the total energy in the first 3 ms of the received pulse, (2) for the first 8.9 ms, (3) for the first 150 ms. The data were normalized to the corresponding values for the source at the aiming point (see Figs. 15 and 16). The geometry of the room is such that there is very little if any reverberation earlier than 3 ms, so energy on this time scale represents reverberationless conditions. However, when the source is away from the aiming point, the acoustic signals from some microphones emerge from the beamformer sooner than others. Far enough away from the aiming point, some microphones even receive no data in the 3 ms window. Because of this, the 3 ms data do not capture all of the direct signal energy and overestimate the array attenuation away from the aiming point. The 8.9 ms energy captures most of the signal pulse, and even at the extremes of source location, almost all microphones contribute some signal to the beamformer output. However, there is also some significant amount of reverberation energy mixed in. The 150 ms data contain all of the direct signal energy and over 90% of the reverberant energy.

Spatial resolution for long or quasicontinuous excitation is measured by the ratio of average power of the excitation to the average power at the array output. All measurements reported here are based on a short chirp excitation and on the integrated energy in the response for various response times. The 150 ms data capture essentially all of the system response. That total energy response ratio will equal the continuous power ratio if the power spectral density of the continuous signal is similar to the magnitude spectrum of the short excitation and if it also has a narrow autocorrelation peak. The spectrum of our test signal covers a broad range from 1500 to over 4000 Hz which includes the range that is important for any subtle speech discrimination. Thus, the 150 ms data are probably a fair representation of our array performance for real-world signals with substantial or important high frequency content.

Section IV presented an idealized mathematical model that neglected background noise, positional measurement errors, reverberation effects, and actual broadband signal spectra. Background noise could be successfully eliminated from the measurements by using the STSS to average enough pulses to make even second-order corrections for noise unnecessary. Errors in the aiming position were compensated by empirically determining where the array was aimed and reckoning change in position from that actual point. Thus, the main problem in comparing calculation to measurement came from the practical impossibility of adding a realistic reverberation model to the simulation. To do a reasonable comparison with measured spatial response required a measurement technique with minimal reverberation content and required attention to the signal spectra in the mathematical model.

The experiment was set up so that no significant reflection would occur for 3 ms, i.e., no reflection added less than about 1 m to the path. This allowed 3 ms or 60 points of simple direct-wave data at each microphone and, by extension, a similar amount of data in the beamformed output. Accumulating power over these first 60 nonzero points from the beamformer gave us a measurement of direct-wave power versus position. With appropriate adaptation, the ideal model can reproduce that data.

To calculate the ideal response to a 3 ms chirp, we used Eq. (7), which is the transfer function for the ideal array. Then, to predict an output from the system, we needed to multiply the transfer function by the DFT of the excitation and take the IDFT of the product. Because the ideal transfer function is identically 1 for the source at the aiming point, we made our calculations using the first 60 points of the measured beamformed output under this condition as the excitation, truncating the data at a zero crossing. The DFT magnitude spectrum of those points appears in Fig. 13. Weights and microphone positions were taken from the data fed to the beamformer. Our source was not a spherical radia-

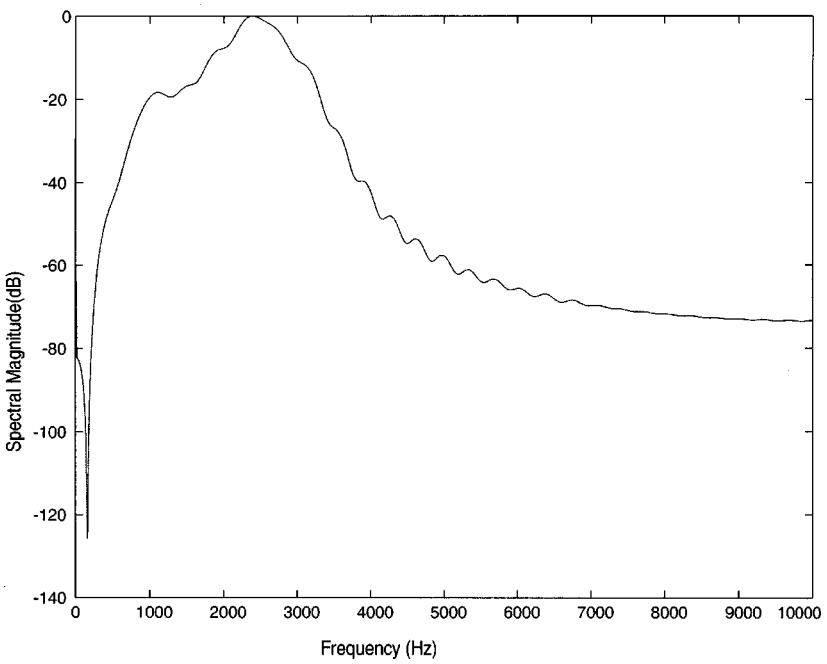


FIG. 13. Magnitude of DFT of the first 3 ms (60 points) of the chirp at the transducer.

tor, as was shown in the measurements of Sec. VI. In particular, there is an attenuation of about 12 dB for the chirp when the microphone is at an angle of 90° away from the front axis of the transducer. The measurements are affected by this source dispersion both by its attenuation of off-axis magnitudes and by any phase differences. We had no reliable data on the phase of the source radiation function and had only limited data on its magnitude as a function of angle from the front of the speaker. The data we did have, however, are an average magnitude for frequencies in the excitation band. These data were presented earlier in Fig. 5. The quantity $H^S(\phi_{ns},\omega)$ in Eq. (7) was approximated as a frequency-independent constant for each microphone by linear interpolation on the dB data underlying Fig. 5.

Some results of this ideal calculation are shown in Fig. 14. The model has been used to compute the beamformed output for three points along the line. The time axis has been chosen so that $t=0$ corresponds to the time relative to the transducer pulse at which a signal from a source at the aiming point will begin to emerge from the beamformer. In Fig. 14(a), at the aiming point, the chirp is 60-points long and starts at time zero. In Fig. 14(b), only 0.25 m away from the aiming point, the response has spread significantly, and in Fig. 14(c), 2 m from the aiming point, the response is about 15 ms long. As expected, the magnitude, when aiming incorrectly, is significantly reduced. Notice that spreading occurs as both advance and retardation since the source gets closer to some microphones and further away from others. The implication for matching this calculation to the measured data is that only the first 60 points of the calculation with significant amplitude match the 3 ms experimental data and many of these points will come before the nominal time zero.

Figures 15 and 16 present the measured spatial resolution based on energy in the three intervals for uniformly weighted and SNR-weighted beamformers, respectively. It also includes the results from the mathematical model.

In comparing the 3 ms measured data to the data from the mathematical model, the shape of the peak and the attenuation about a meter from the aiming point match remarkably well. Inclusion of the radiation pattern function in the calculation significantly improves the match because away from the aiming point microphones far off the source axis are the largest contributors to the output. Attenuation at such angles is almost 10 dB and failure to take that into account leads to about 8 dB greater discrepancy with the measured data. There is a 3–5 dB difference in attenuation about 0.25 m to the right of the peak that is poorly understood.

While the agreement between model and experiment for the 3 ms data confirms the adequacy of the model and its parameters, it does not represent the actual attenuation of the array even for just direct-path signals. As is clear from Fig. 14(c), deliberately limiting power accumulation to 3 ms may underestimate the total output energy by as much as a factor of 3. The direct-wave asymptotic attenuation is, therefore, from 3 to 5 dB higher than the 3 ms data. This is similar to the observed energy in the 8.9 ms interval, which captures at least 70% of the direct wave energy but has a little reverberation energy mixed in. Apparently the 8.9 ms data are close to what one would observe for the whole system in anechoic conditions. The effect of reverberation is clear from the difference between the curves for the 8.9 and 150 ms accumulations. There is an approximately 3 dB decrease in the asymptotic attenuation over a wide range of source location. In our environment the total reverberant energy is about the same as the direct signal energy (cf. Fig. 10, which shows this ratio varying from $+7$ to -7 dB with the mean around 0 dB). With the source significantly away from the aiming point, the system essentially adds signal energies with no gain from correlation. The reverberation signals correlate poorly between channels and with the direct signals and so simply double the net received energy, accounting for the observed increase of 3 dB.

The effect of the choice of weighting coefficients is shown in Fig. 17, which shows the 3 ms data for the two

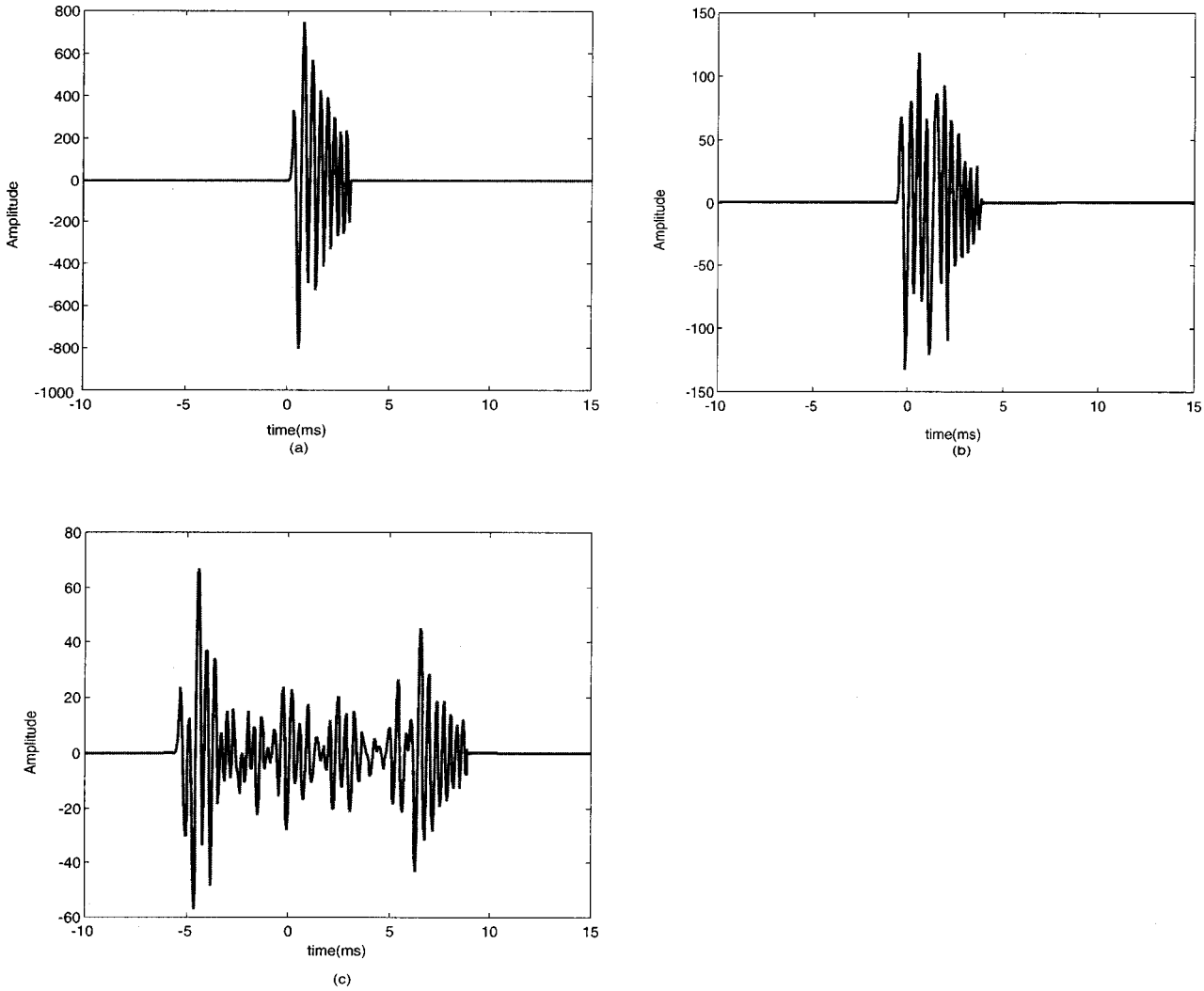


FIG. 14. Ideal response driven by 3 ms chirp at (a) the aiming point, (b) 0.25 m from the aiming point, and (c) 2 m from the aiming point.

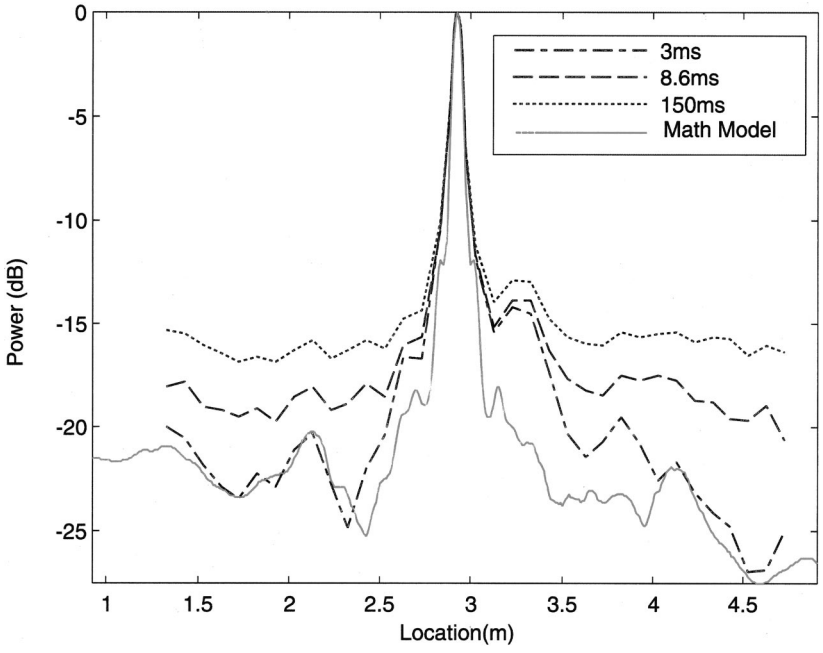


FIG. 15. Measured and ideal responses for the array over a line parallel to the *Y* axis and through the source.

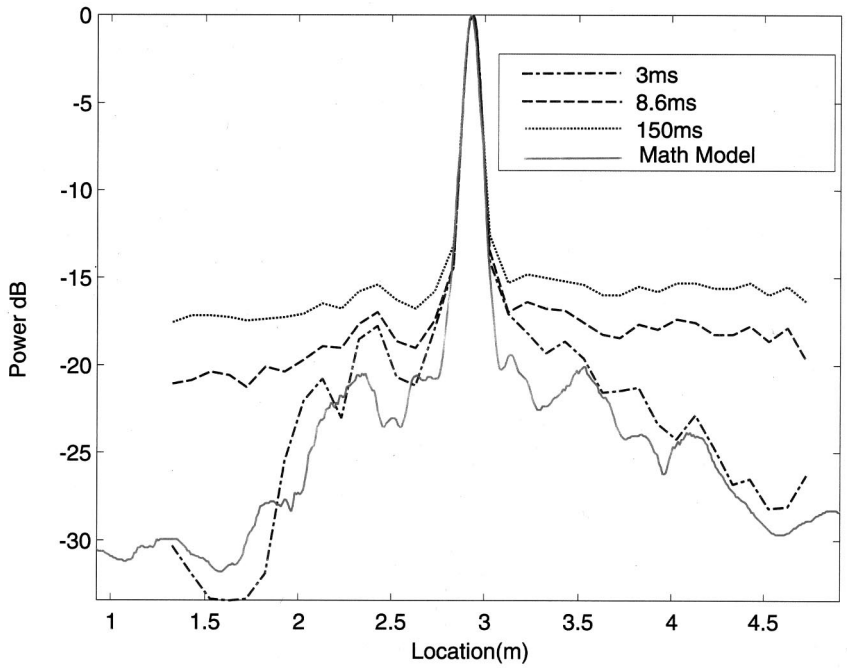


FIG. 16. Measured and ideal, S/N^2-weighted responses for the array over a line parallel to the *Y* axis and through the source.

weight sets plotted against an expanded *Y* scale. Uniform weighting gives a slightly narrower peak than SNR weighting, which is expected since the latter reduces the effective aperture somewhat. (The weight sets are described in more detail with the SNR measurements.) The asymptotic suppression is similar for the two cases.

B. Performance at the aiming point

The two sets of weights were intended to explore optimization of the signal-to-noise ratio at the aiming point. One set provided equal weighting of all channels regardless of the channel SNR. The other set weighted channels by the ratio of signal amplitude to the mean-square-noise amplitude. This is known to be an optimum SNR weighting for a system with uncorrelated noise and no reverberation.[32] In calculating the SNR weights, the relative signal amplitude was determined from the first 3 ms of the individual channel responses. Per-channel noise energy was calculated by averaging several single-shot (unaveraged) measurements with the source turned off. Figure 18 shows the resulting weights normalized to the highest value. Note that the microphones in panel A (see Fig. 1) contribute less than their counterparts in panel H because of the high level of fan noise from the HMA console that is near the former set.

To calculate the measured SNRs, we computed the total energy in the first 8.9 ms of the beamformed data as the direct signal energy. The noise was estimated by averaging the energy in several 8.9 ms measurements of the beamformer output with the source turned off. The reverberant energy was extracted from the averaged beamformed data by

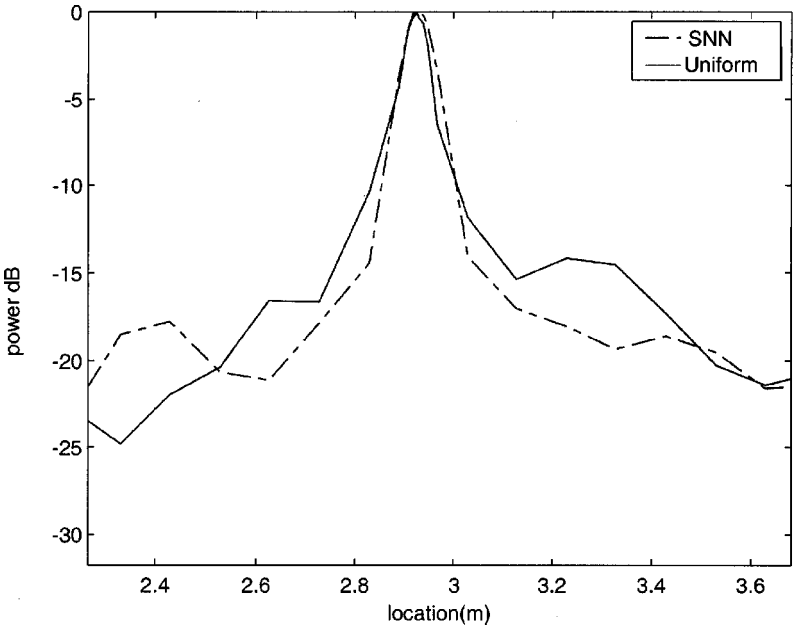


FIG. 17. Main lobe region for S/N^2 and uniformly weighted measured responses using only the direct wave.

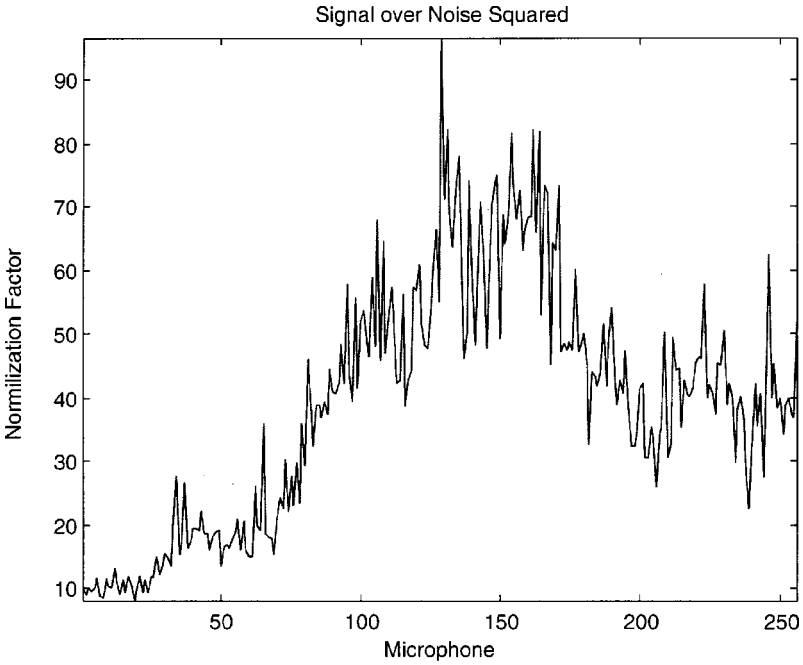


FIG. 18. Relative weights for S/N^2 weighting.

computing the total energy from 9 to 300 ms. This ignores the reverberant energy during the last 6 ms of direct pulse time. Since the decay of reverberation energy in the room is approximately exponential (Fig. 7) with a time constant of about 61 ms, it is possible to make small corrections to the signal and reverberant energy calculations if one assumes that reflections begin approximately 6 ms after the pulse. Corrections are on the order of 10% and the residual errors from the overlap of signal and reverberation are of the same order. Table I summarizes the measured and calculated SNRs. For SNR-weighted beamformers with perfectly correlated signals and uncorrelated noise, the SNR of the output is given by

$$\mathrm{SNR}=\sum_i \frac{\overline{a_i^2}}{\overline{N_i^2}}, \tag{10}$$

where $\overline{a_i^2}$ is the mean square signal in microphone i and $\overline{N_i^2}$ is the mean-squared noise. With similar assumptions, an equiweighted beamformer has a SNR given by

$$\mathrm{SNR}=\frac{(\Sigma a_i)^2}{\Sigma N_i^2}, \tag{11}$$

where a_i is the rms signal energy in channel i. To provide a theoretical comparison to the measured data for the array, we calculated these SNRs from data for a set of single channel measurements. In using the same SNR models for both background noise and reverberation there is an implicit assumption that reverberation signals do not correlate between channels very well.

The gain in signal-to-noise ratio for the array output over the best single microphone is about 10 dB for reverberant energy and 6 dB for background noise. If all microphones had the same starting SNR, the theoretical limiting gain would be $\sqrt{N}$ or 24.1 dB. As the tabular data (Table I) indicate, the real signal quality limits the theoretical gain to about 18 dB. The observed gain is about 7–12 dB lower. This must be due to some combination of correlation of noise and reverberation between channels and the imperfect correlation of signals between channels. Measurements of the correlation coefficient between the signal from directly in front of the source to measurements around the source suggest that the two factors are of similar importance. The background noise is primarily acoustic (computer fan noise) from a location near one section of the array. The lower effectiveness of the array for noise may be due to the noise correlation caused by this geometry. Both theory and observation show a relatively modest gain of about 2 dB for optimal SNR weighting constants over equal weighting, particularly if the source is not close to the array. This suggests that optimal weighting is only useful if the computation required to

TABLE I. Comparison of signal-to-noise ratios with different weightings.

Configuration	Signal-to-reverberant-noise ratio (dB)	Signal-to-background-noise ratio (dB)	Signal-to-total-noise ratio (dB)
Best single microphone	8.0	26.3	7.9
Equiweighted beamformer	17.6	31.4	17.5
SNR-weighted beamformer	18.6	33.0	18.4
Theoretical comparison for equiweighted	24.6	40.8	24.4
Theoretical comparison for SNR weighted	25.3	42.6	25.2

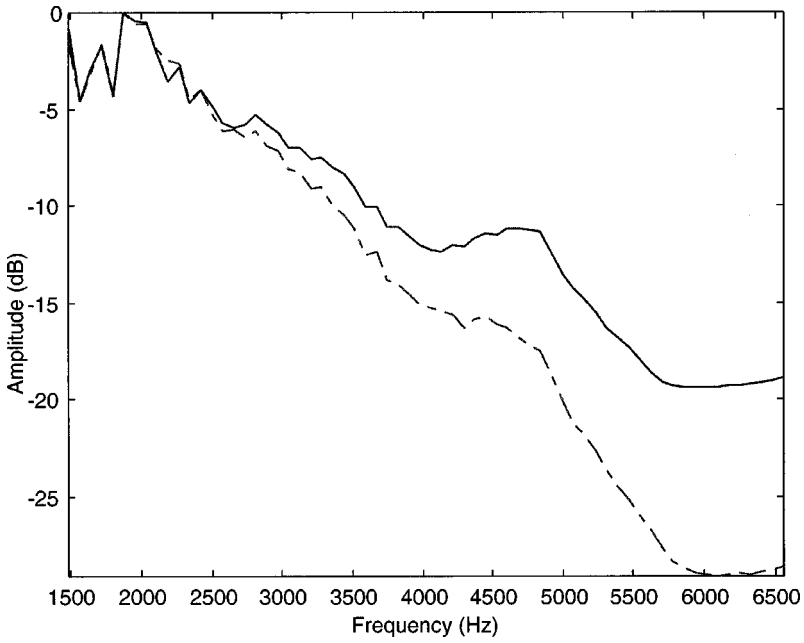


FIG. 19. Relative frequency response of equiweighted (dashed curve) and SNR-weighted (solid curve) beamformers.

implement it is very low cost or if the source gets very close to one part of the array.

The frequency response for an ideal system at the aiming point is perfectly flat. Reverberation introduces fairly high-Q resonances in the room that make equilibrium sinusoidal measurements of frequency response quite problematic. To find an approximate frequency response of the actual system over a limited range of frequencies, we compared the first 8.9 ms of the beamformed signal to a measurement with a single microphone directly in front of the speaker. (This latter measurement was made during the speaker characterization described in Sec. VI under conditions that guaranteed no reverberation energy within the pulse duration.) The excitation was the 8.9 ms chirp (2000 Hz–6500 Hz) discussed in the following. The ratios of the magnitude spectra for equiweighted and SNR-weighted beamformers over 1500–6500 Hz to the forward reference spectrum are shown in Fig. 19.

The peak response in both cases is at 2000 Hz. The relatively lower response at higher frequencies is largely due to the source radiation pattern, which shows reduced emission of high frequencies at angles away from the front of the source. At 2000 Hz, essentially all microphones contribute to the output. As frequency increases, progressively fewer microphones contribute because the radiation pattern narrows and they no longer receive significant signal energy. The curve for the equiweighted beamformer shows more high-frequency attenuation because the microphones at larger angles are weighted more than in the SNR-weighted beamformer. As these are more affected by dispersion, their loss has greater effect.

IX. SUMMARY AND CONCLUSIONS

In this first effort to understand the factors contributing to the performance of a large-aperture system in a real environment, we have successfully related results from an idealized, mathematical model to real measurements. A novel test source that provides precisely synchronized and highly repeatable acoustic signals allowed us to separate the direct-path signal from the effects of reverberation and background noise. The direct-signal measurements reasonably match the predicted responses.

Our 256 microphone array using delay-and-sum beamforming shows broadband spatial discrimination in the form of a single central peak about 0.4 m in diameter at its 3 dB points. Asymptotic attenuation away from its aiming point is about 15 dB. A simulation model suggests that attenuation of the direct wave signals is about 18 dB. Reverberation energy in our environment is comparable to the direct signal energy and so approximately doubles the asymptotic output energy, decreasing the observed attenuation by 3 dB.

The signal-to-noise gain from using the array is about 10 dB for reverberation and 6 dB for background noise. Theoretical calculations show a limit of about 17 dB for both these cases. Interchannel correlation properties of the signal and noise components account for the differences (see Sec. VIII B). Within the constraints of a delay-and-sum system, only modest improvements are possible with optimal weighting, so calculating such weights is only practical with methods having a low computational cost.

The effect of source dispersion, as expressed in the radiation pattern function $H^S(\theta,\phi,\omega)$, is significant in several ways. Some of the observed shortfall in expected SNR is due to imperfect correlation of signals emitted at different polar angles. A good dispersion model and a knowledge of source orientation could recover some of this loss. Second, since one needs position estimates to do beamforming, an orientation estimate is the only additional requirement for low-cost calculations of SNR optimized weights. Third, source dispersion generally imposes a non-negligible frequency response even for a source at the aiming point. Finally, it is a subtle but important point that the room transfer function (RTF) depends on source orientation through $H^S(\theta,\phi,\omega)$. Any beamformer that exploits the RTF to improve SNR will have to take careful account of this effect.

For our rather typical environment, the reverberation energy is comparable to the direct signal energy. This suggests that a strategy making use of that energy rather than simply trying to suppress it would improve the array performance. Such techniques include matched filtering[17] and approximate inverse filtering.[20] However, the careful measurements of RTF reported here show the complexity of this function and highlight its dependence on source orientation. The implication is that exploiting the RTF will be quite difficult in unconstrained conditions.

The observations reported here have explained a number of phenomena we have seen while developing algorithms. With renewed confidence in the computational model, we are moving to increase the system to 512 microphones in an array fully surrounding the room. Earlier work using the same theory, presented at ICASSP-99,[7] showed this to be an optimal configuration. We expect the new array will show improved spatial resolution in the *x* direction and more uniform response to sources that rotate as well as translate.

[1] J. L. Flanagan and H. F. Silverman, ''Material for international workshop on microphone-array systems: Theory and practice,'' LEMS Technical Report No. 113, LEMS, Division of Engineering, Brown University, Providence, RI, October 1992.

[2] J. L. Flanagan and H. F. Silverman, ''Material for international workshop on microphone-array systems: Theory and practice,'' Technical report, CAIP, Rutgers University, Piscataway, NJ, October 1994.

[3] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, ''A localization error criterion for microphone-array sensor placement,'' Proceedings of ICASSP-1996, Atlanta, GA, May 1996, pp. 901–904.

[4] M. S. Brandstein, J. L. Flanagan, and H. F. Silverman, ''Material for fourth roundtable of microphone-array technology,'' Technical Report No. 151, Harvard University, Cambridge, MA, October 2000.

[5] H. F. Silverman, ''Some analysis of microphone arrays for speech data acquisition,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 1699–1712 (1987).

[6] M. Berger and H. F. Silverman, ''Microphone array optimization by stochastic region contraction (SRC),'' IEEE Trans. Signal Process. **39**, 2377–2386 (1991).

[7] H. F. Silverman and W. R. Patterson III, ''Visualizing the performance of large-aperture microphone arrays,'' Proceedings of ICASSP-1999, Phoenix, AZ, March 1999, pp. 969–972.

[8] J. L. Flanagan, A. C. Surendran, and M. Rahim, ''Microphone array preprocessing of speech,'' Proceedings of DOD Speech Research Symposium XII, Piscataway NJ, June 1992.

[9] T. B. Hughes, H-S Kim, J. H. DiBiase, and H. F. Silverman, ''Using a real-time, tracking microphone array as input to an hmm speech recognizer,'' Proceedings of ICASSP-98, Seattle, WA, May 1998, pp. I-249, I-252.

[10] K. Kiyohara, Y. Kaneda, S. Takahashi, H. Nomura, and J. Kojima, ''A microphone-array system for speech recognition,'' Proceedings of ICASSP-97, Munich, Germany, 22–25 April 1997, pp. I-215, I-218.

[11] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, ''A localization-error-based method for microphone-array design,'' Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Atlanta, GA, 7–10 May 1996, pp. II-901–II-905.

[12] Y. Grenier, ''A microphone array for car environments,'' Proceedings of ICASSP92, San Francisco, CA, April 1992, pp. I-308–I-309.

[13] D. Van Compernolle, W. Ma, F. Xie, and M. Van Diest, ''Speech recognition in noisy environments with the aid of microphone arrays,'' Speech Commun. **9**, 433–442 (1990).

[14] H. F. Silverman, W. R. Patterson III, and J. M. Sachar, ''First measurements of a large-aperture microphone array system for remote audio acquisition,'' Proceedings of IEEE International Conference on Multimedia and Expo, New York, July/August 2000, Session TP0, Paper 5.

[15] H. F. Silverman, W. R. Patterson III, and J. L. Flanagan, ''The huge microphone array (HMA). II,'' IEEE Trans. Concurrency **7**, 32–47 (1999).

[16] H. F. Silverman, W. R. Patterson III, and J. M. Sachar, ''Early results for a large-aperture microphone array system,'' Proceedings of SAM2000, Boston, MA, March 1999, pp. 207–211.

[17] E. Jan, P. Svaizer, and J. Flanagan, ''Matched-filter processing of microphone array for spatial volume selectivity,'' Proceedings of ISCAS95, IEEE, 1995, pp. 1460–1463.

[18] R. Renomeron, D. Rabinkin, J. French, and J. Flanagan, ''Small-scale matched filter array processing for spatially selective sound capture,'' J. Acoust. Soc. Am. **102**, 3208 (1997).

[19] D. Rabinkin, D. Macomber, R. Remomeron, and J. Flanagan, ''Optimal truncation time for matched filter array processing,'' Proceedings of ICASSP-98, Seattle, WA, May 1998, pp. VI-3629–VI-3632.

[20] J. L. Flanagan, Q. Lin, R. E. Van Dyck, J. Upton, J. French, E. E. Jan, D. Rabinkin, and A. C. Surendran, ''A large-scale, intelligent three-dimensional microphone array sound capture system,'' Technical report, CAIP, Rutgers University, Piscataway, NJ, January 1995.

[21] O. Hoshuyama and A. Sugiyama, ''A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters,'' Proceedings of ICASSP-96, Atlanta, GA, 7–10 May 1996, pp. II-925–II-932.

[22] H. F. Silverman, W. R. Patterson III, and J. L. Flanagan, ''The huge microphone array (HMA). I,'' IEEE Trans. Concurrency **6**, 36–46 (1998).

[23] H. F. Silverman, W. R. Patterson III, and J. M. Sachar, ''An experiment that validates theory with measurements for a large-aperture microphone array,'' Proceedings of ICASSP-2001, Salt Lake City, UT, May 2001, p. SAM P7.3.

[24] R. C. Heyser, ''Acoustical measurements by time-delay spectrometry,'' J. Audio Eng. Soc. **15**, 370–382 (1967).

[25] C. R. Cable and J. K. Hilliard, ''The paractical application of time-delay spectrometry in the field,'' J. Audio Eng. Soc. **28**, 302–309 (1980).

[26] G. Stanley, ''TDS computing,'' Audiology, 38–42 (1983).

[27] D. Davis, ''Tds in sound measurements,'' Audiology, 46–52 (1978).

[28] D. Davis and C. Davis, ''Tds—a progress report,'' Audiology, 58–64 (1979).

[29] K. J. Jebelian, ''A real-time digital approach to time-delay spectrometry,'' Audio Engineering Society Preprint No. 2988, presented at the 89th AES Convention, Los Angeles, CA, September 1990.

[30] P. Meuse, ''The radiation pattern of speech production: Implications for microphone array systems,'' Ph.D. thesis, Brown University, Providence, RI, May 2000.

[31] M. Rossi, *Acoustics and Electroacoustics* (Artech, Norwood, MA, 1988).

[32] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques* (Prentice–Hall, Englewood Cliffs, NJ, 1993).

[33] ''Parts express electronics and more,'' Springboro, OH, Parts Express Catalog, 91st ed., 2000.

An acoustic-logging transmission-network model

Lin Fa and John P. Castagna
Sarkeys Energy Center, 100 East Boyd Street, Norman, Oklahoma 73019

Jens M. Hovem
The Norwegian University of Science and Technology, N-7491 Trondheim, Norway

Daqun Dong
Marine Institute, Northwest Polytechnic University, Xi'an, Shaanxi 710071, People's Republic of China

(Received 29 December 2000; accepted for publication 10 January 2002)

The acoustic logging process can be described as a signal transmission system with a corresponding transmission network model. This model gives the relations between driving-voltage signal, the electrical-acoustic conversion at the source, acoustic properties of the propagation media (borehole fluid and formation around borehole), the pressure response in the borehole, the acoustic-electrical conversion at the receiver and recorded logging signal. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1456925]

PACS numbers: 43.58.Vb, 43.58.Wc, 43.20.Bi, 43.20.Mv [SLE]

I. INTRODUCTION

Acoustic logging of boreholes is widely used to (1) evaluate rock properties and porosity of formation around borehole, (2) identify fracture, (3) provide images of open-hole wall and casing corrosion, (4) appraise the cement bond quality of cased-well, (5) measure the diameters of both borehole and cased-well, and (6) calibrate seismic data. Typically, some assumed source functions are used to compute synthetic microseismograms. Examples in the literature include the Ricker wavelet,[1] Tsang wavelet,[2] or Gaussian impulse wavelet.[3] Different kinds of acoustic transducer are used as source and receiver in acoustic logging tools. These include monopole and dipole transducers and thin cylindrical transducers polarized in radial and tangential directions with different resonance frequencies. Cheng and Toksöz[4] found that the amplitude and frequency variations of Stoneley waves in synthetic microseismogram was more enhanced compared with actual borehole acoustic logging signals, and hypothesized that this may be because the frequency responses of the logging tool sources and receivers are narrower than the ones used in their calculations. The source spectrum of the acoustic signal has a strong effect on the excitation of pseudo-Rayleigh waves. In order to understand the acoustic signal spectrum, it is necessary to consider driving-voltage signal, electrical–acoustic conversion at the source, and acoustic–electrical conversion at the receiver. Berlincourt, Curran, and Jaffe[5] gave the equivalent circuit of the piezoelectric shell transducer. In order to calibrate a transducer and separate the interrogating radiation from the scattered or reflected radiation in a test facility of limited size, Piquette presented a method for the transient suppression of both a spherical transducer[6,7] and various transducer types.[8] Piquette and Forsthe[9] put forward generalized methods for determining the required wave shape for the driving-voltage wave form that produces no transients in the acoustic field radiated from an electroacoustic projector. Holly[10] described a method for the generation of broadband acoustic transmissions. Fa, Lin, and Chen[11] derived the source functions of a cylindrical transducer for several driving-voltage signals and performed the analysis in the time and frequency domains. In this paper, we derive the electrical–acoustic and acoustic–electrical transmission functions of the piezoelectric transducer, considering the effects of the output impedance of the driving circuit, the input impedance of the measurement circuit and the complex mechanical load of the surrounding coupling medium around the transducer. According to the electrical–acoustic and acoustic–electrical transmission functions of the transducer, by choosing suitable driving-voltage signal, the electrical parameters of the driving and measurement circuits, the physical parameters of the coupling medium and the physical and geometric parameters of transducer, we can make the transducer radiate the desired acoustic signal to optimize the measurement signal for various applications. We also utilize the concept of information and signal transmission to analyze the acoustic logging process, describe the acoustic logging process as a signal transmission system, and derive its transmission network model. This model consists of three submodels: a model for the source, a model for the propagation media (borehole fluid and formation around borehole) and a model for the receiver. In this model (1) the acoustic source acts like an electrical–acoustic filter, (2) the receiver is equivalent to an acoustical–electric low-pass filter, (3) the electric terminals of both the source and the receiver are the input and output terminals of the transmission network and, (4) the driving-voltage signal and the measured logging signal are the input and output signals of the network. Since the input signal of the network (driving-voltage signal), the properties of the source and receiver, the physical parameters of the borehole fluid and the borehole size can be assumed to be known, one can potentially invert for the formation properties from the output signal of the network (the recorded logging signal).

II. FORMULATION

In the following, the case of the medium around borehole being an infinite elastic body is discussed.

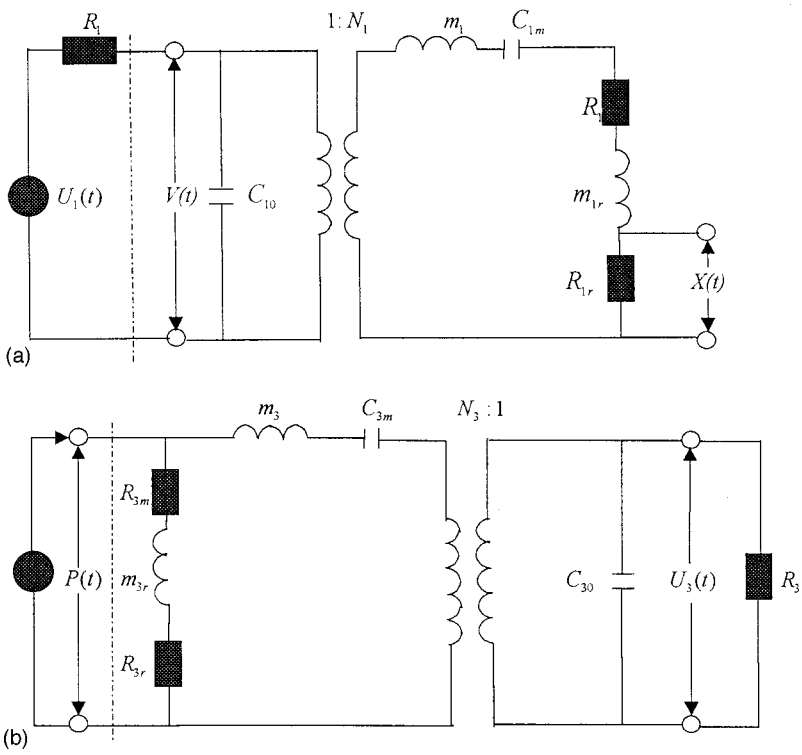


FIG. 1. Equivalent circuits of the spherical shell transducer. (a) Source. (b) Receiver.

For a piezoelectric thin spherical shell transducer with average radius r_b and wall thickness l_t, polarized in the radial direction with electrodes connected to the inner and outer surfaces, its electric–acoustic and acoustic–electric equivalents (source and receiver)[5,6,11–13] are shown in Figs. 1(a) and (b). The description and expression of the symbols in Fig. 1 are shown in Table I.

TABLE I. Description and expression of symbols in Fig. 1. ρ_p is the density of the transducer material; s_{i11}^E, s_{i12}^E, d_{i31}, and ε_{i11}^T are the strain, piezoelectric, and dielectric constants of the transducer material, respectively, $S_{iC}^E=(s_{i11}^E+s_{i12}^E)/2$, $k_{i31}=d_{i31}/\sqrt{S_{iC}^E\varepsilon_{i33}^T}$, and $k_m=\omega_1/\nu_m$; the subscript $i=1$ stands for source and $i=3$ stands for receiver, where, $s_{111}^E=s_{311}^E=s_{11}^E$, $s_{112}^E=s_{312}^E=s_{12}^E$, $d_{131}=d_{331}=d_{31}$, $\varepsilon_{111}^T=\varepsilon_{311}^T=\varepsilon_{11}^T$, $k_{131}=k_{331}=k_{31}$, $S_{1C}^E=S_{3C}^E=S_C^E$, $l_{1t}=l_{3t}=l_t$, $r_{1b}=r_{3b}=r_b$, and ν_m is the acoustic velocity of the coupling fluid around the transducer.

Symbol	Description	Expression
R_1	Output resistance of driving circuit	
$U_1(t)$	Driving-voltage signal	
$V(t)$	Voltage signal of electric terminals of source	
$P(t)$	Pressure response signal in borehole	
$X(t)$	Acoustic pressure output signal of source	
R_3	Input resistance of measurement circuit	
$U_3(t)$	Electric output signal of receiver	
C_{i0}	Clamped capacitance of transducer	$4\pi r_{ib}^2\varepsilon_{i33}(1-k_{i31}^2)/l_{it}$
N_i	Mechanical–electrical conversion coefficient of transducer	$4\pi r_{ib}d_{i31}/S_{iC}^E$
m_i	Mass of transducer	$4\pi r_{ib}^2 l_{it}\rho_p$
C_{im}	Elastic stiffness of transducer	$s_{iC}^E/4\pi l_{it}$
R_{ir}	Radiation force resistance of transducer	$4\pi k^2 r_{ib}^3\rho_m\nu_m/(1+k_m^2r_{ib}^2)$
m_{ir}	Ratio of radiation impedance to resonance frequency	$4\pi\rho_m r_{ib}^2/(1+k_m^2r_{ib}^2)$
R_{im}	Friction force resistance	

According to Fig. 1(a), the electrical–acoustic transmission function of the source in the time domain can be written as follows:

$$h_1(t)=K_1\exp[-\beta_1 t]+K_2\exp[-\alpha_1 t]\cos(\omega_1 t+\theta_1), \tag{1a}$$

and its resonance frequency is

$$f_1=\omega_1/2\pi=\sqrt{3}(x_1-y_1)/4\pi, \tag{1b}$$

where $\tilde{M}_1=m_1+m_{1r}$, $\tilde{R}_1=R_{1m}+R_{1r}$, $a_1=(R_1C_{10}\tilde{R}_1C_{1m}+C_{1m}\tilde{M}_1)/(R_1C_{10}C_{1m}\tilde{M}_1)$, $b_1=[R_{10}C_{1m}+N_1^2(R_1C_{10}+\tilde{R}_1C_{1m})]/(N_1^2R_1C_{10}C_{1m}\tilde{M}_1)$, $c_1=1/(R_1C_{10}C_{1m}\tilde{M}_1)$, $d_1=R_1C_{1m}/(R_1C_{10}C_{1m}\tilde{M}_1)$, $e_1=d_1R_{1r}C_{1m}$, $p_1=b_1-a_1^2/3$, $q_1=c_1+2a_1^3/27-a_1b_1/3$, $\tilde{D}_1=(p_1/3)^3+(q_1/2)^2$, $x_1=(-q_1/2+\sqrt{\tilde{D}_1})^{1/3}$, $y_1=(-q_1/2-\sqrt{\tilde{D}_1})^{1/3}$, $\alpha_1=a_1/3-(x_1+y_1)$, $\beta_1=(x_1+y_1)/2+a_1/3$, $A_1=\beta_1-\alpha_1$, $B_1=d_1(\beta_1-A_1)/[2(\omega_1^2+A_1^2)]$, $C_1=-d_1(\omega_1^2+\beta_1A_1)/[2\omega_1(\omega_1^2+A_1^2)]$, $\theta_1=\mathrm{arctg}(C_1/B_1)$, $K_1=-N_1d_1\alpha_1/(A_1^2+\omega_1^2)$, $L_1=(B_1^2+C_1^2)^{1/2}$, and $K_2=L_1N_1$.

From Fig. 1(b), the acoustic–electrical transmission function of the receiver in the time domain and its resonance frequency can be written as

$$h_3(t)=K_3\exp[-\alpha_3 t]+2K_4\exp[-\beta_3 t]\cos(\omega_3 t+\theta_3), \tag{2a}$$

$$f_3=\omega_3/2\pi=\sqrt{3}(x_3-y_3)/4\pi, \tag{2b}$$

where $\tilde{M}_3=m_3+m_{3r}$, $\tilde{R}_3=R_{3m}+R_{3r}$, $a_3=(C_{30}R_3C_{3m}\tilde{R}_3+C_{3m}\tilde{M}_3)/C_{30}R_3C_{3m}\tilde{M}_3$, $b_3=(C_{30}R_3+C_{3m}\tilde{R}_3+N_3^2R_3\times C_{3m})/(C_{30}R_3C_{3m}\tilde{M}_3)$, $d_3=N_3R_3C_{3m}/(C_{30}R_3C_{3m}\tilde{M}_3)$, $e_3=d_3R_{3r}C_{3m}$, $c_3=1/(C_{30}R_3C_{3m}\tilde{M}_3)$, $p_3=b_3-a_3^2/3$, $q_3=c_3+2a_3^3/27-a_3b_3/3$, $\tilde{D}_3=(p_3/3)^3+(q_3/2)^2$, $x_3=(-q_3/2+\sqrt{\tilde{D}_3})^{1/3}$, $y_3=(-q_3/2-\sqrt{\tilde{D}_3})^{1/3}$, $\alpha_3=a_3/3-(x_3+y_3)$,

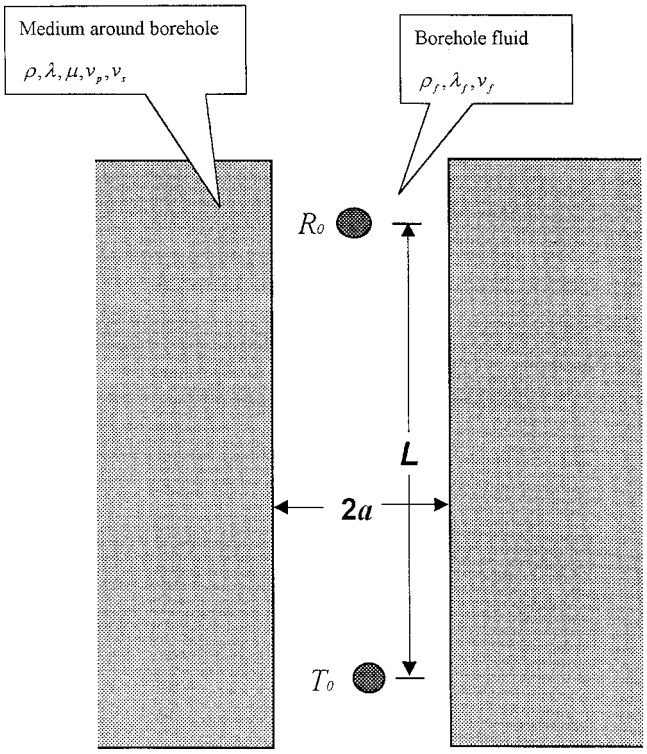


FIG. 2. Geometrical configuration of acoustic logging.

$\beta_3=(x_3+y_3)/2+a_3/3$, $A_3=\beta_3-\alpha_3$, $B_3=\omega_3^2+A_3^2$, $C_3=N_3R_3C_{m3}/2a_3\omega_3$, $E_3=-C_3(\omega_3^2+\beta_3A_3)$, $D_3=C_3\omega_3(\beta_3-A_3)/B_3$, $K_3=-d_3\beta_3/(A_3^2+\omega_3^2)$, $K_4=(D_3^2+E_3^2)^{1/2}$, and $\theta_3=\text{arctg}(D_3/E_3)$.

As shown in Fig. 2, the logging tool is set in a fluid-filled cylindrical borehole of radius a, which embeds in an infinite elastic medium. T_0 is the acoustic source, which is located at the origin; R_0 is the receiver, which is located at any point z on the borehole axis; ρ_f, λ_f, and ν_f are the density, Lamb's coefficient, and acoustic velocity of the borehole fluid; ρ, λ, μ, ν_p, and ν_s are the density, Lamb's coefficients, and the *P*- and *S*-wave velocities of the formation around borehole. In the far field, when T_0 transmits acoustic signals outwards, the effect of the propagation media (formation around borehole and borehole fluid) on the transmitted acoustic signals can be described by

$$h_{2il}(t,z)=\int_{-\infty}^{+\infty}H_2(\omega)\exp[j\omega t]d\omega \tag{3a}$$

and

$$H_{2il}(\omega,z)=H_d(\omega,z)+H_r(\omega,z)=\frac{\exp[j\omega z/V_f]}{z}+\frac{1}{(2\pi)^2}\int_{-\infty}^{+\infty}A(k_z,\omega)\exp[jk_zz]dk_z, \tag{3b}$$

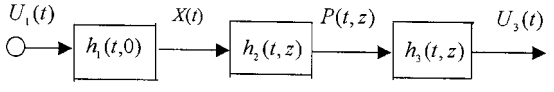


FIG. 3. Acoustic logging transmission network in time domain.

TABLE II. Physical parameters of piezoelectric material *PZT-7A* and acoustic impedance of coupling fluid.

$\rho_p\times10^3$ (kg/m^3)	$d_{31}\times10^{-12}$ (m/V)	$\varepsilon_{33}^T\times10^{-9}$ (F/m)	$s_{12}^E\times10^{-12}$ (m^2/N)	$s_{11}^E\times10^{-12}$ (m^2/N)	$Z_m\times10^6$ [kg/(m$^2\cdot$s)]
7.6	−60	3.7613	−3.2	10.7	1.2205

where

$$A=-\frac{F_1k_raH_1^{(1)}(k_ra)+F_2H_0^{(1)}(k_ra)}{F_1k_raJ_1(k_ra)+F_2J_0(k_ra)}, \tag{3c}$$

$$F_1=-\mu\left[2\frac{k_s^2}{a^2}-(k_s^2-2k^2)^2\frac{H_0^{(1)}(k_r^{(c)}a)}{k_r^{(c)}aH_1^{(1)}(k_r^{(c)}a)}-4k^2k_r^{(s)}\frac{H_0^{(1)}(k_r^{(s)}a)}{aH_1^{(1)}(k_r^{(s)}a)}\right], \tag{3d}$$

$$F_2=\lambda k^2k_s^2, \tag{3e}$$

ω is the angular frequency; the borehole fluid wave number $k=\omega/\nu_f$; k_z is axial wave number and its radial component $k_r=\sqrt{k^2-k_z^2}$; $J_0(k_rr)$ is the zeroth-order Bessel function; $k_r^{(c)}=(k_c^2-k^2)^{1/2}$; $k_r^{(s)}=(k_s^2-k^2)^{1/2}$; $k_p^2=\omega^2/\nu_p^2$; and $k_s^2=\omega^2/\nu_s^2$.

From Fig. 2 and the above relations, the transmission network of acoustic logging can be obtained as shown in Fig. 3. When T_0 is excited by the driving-voltage signal $U_1(t)$, the pressure response in the borehole at the position of R_0 and the electric signal output by R_0 (i.e., the recorded logging signal) can be expressed by Eqs. (4) and (5),

$$P(t,z)=U_1(t)*h_1(t,0)*h_2(t,z), \tag{4}$$

$$U_3(t,z)=P(t,z)*h_3(t,z). \tag{5}$$

III. SIMULATION AND DISCUSSION

In the following calculation, let the thin spherical shell transducer consist of piezoelectric material *PZT-7A*, its average radius $r_b=8$ cm and wall thickness $l_t=0.8$ cm, respectively, the coupling fluid be transformer oil, the friction force resistance $R_m=0.8\pi r_b^2Z_m$, the output impedance of the driving circuit $R_1=50\Omega$ and the input impedance of the measurement circuit $R_3=50\,\text{k}\Omega$. The physical parameters of *PZT-7A* and the specific acoustic impedance Z_m of the coupling fluid are shown in Table II.[11]

A. Electrical–acoustic conversion properties of the transducer

According to Eq. (1a), the calculated electrical–acoustic transmission property of the source is shown in Figs. 4(a) and (b). The source has the electrical–acoustic transmission property of negative original-phase, its resonance frequency f_1 is 11.268 kHz and it acts like an electrical–acoustic band-pass filter.

B. Relation between driving-voltage signal and acoustic signal radiated by source

In order to make the acoustic source radiate the desired acoustic signal, according to Eqs. (1a) and (1b), we use three

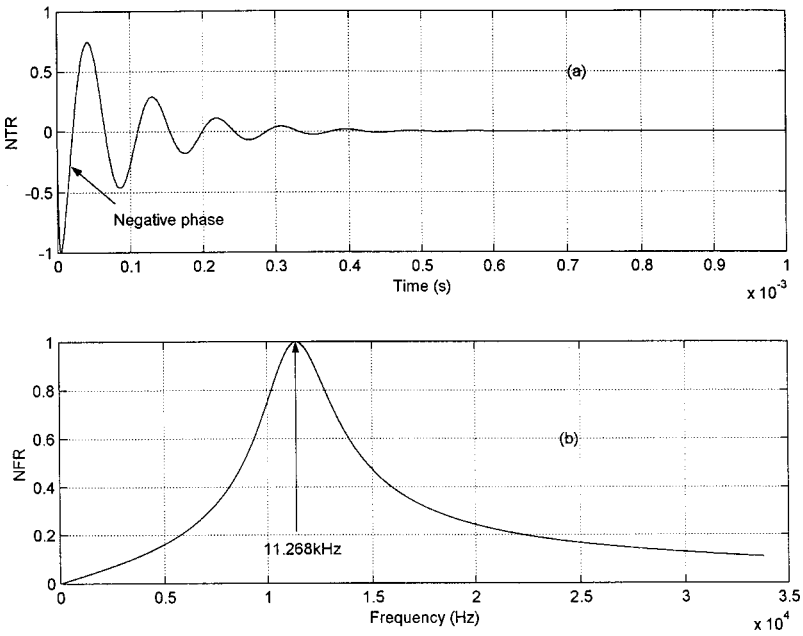


FIG. 4. Electric–acoustic transmission property of source. NTR is the normalized amplitude of electrical–acoustic transmission property of source and NFR is the normalized amplitude spectrum of electrical–acoustic transmission property of source. (a) Time domain. (b) Frequency domain.

kinds of driving-voltage signals: (1) a gated sinusoidal, (2) an impulse, and (3) a narrow boxcar function for computing. The chosen parameters are shown in Table III. The expressions of the above three kinds of driving-voltage signals in the time domain are, respectively, as follows:

$$U_1(t)=[H(t)-H(t-t_0)]U_0\sin(\omega_1 t), \tag{6}$$

$$U_1(t)=U_0 t^3\exp[-\delta t], \tag{7}$$

$$U_1(t)=[H(t)-H(t-t_1)]U_0, \tag{8}$$

where U_0 is the amplitude constant of the driving-voltage signal; δ is the damping coefficient of the impulsive driving-voltage signal; t_0 is the window width of the gated sinusoidal driving-voltage signal; t_1 is the width of the narrow boxcar function driving-voltage signal; $H(t)$ and $H(t-t_i)$ $(i=0$ or $1)$ are Heaviside unit step functions.

When the source is excited by the three kinds of driving-voltage signals, the calculated wave forms (RAWF) and amplitude spectra (RAAS) for the radiated acoustic signals are shown in Figs. 5 and 6. The first half-wave of the gated sinusoidal driving-voltage signal, the impulsive and a narrow boxcar function driving-voltage signals are positive, while the first half waves of the radiated acoustic signals are negative. This phenomenon is caused by the electric-acoustic transmission property of the negative original phase of the source.

When the source is excited by the gated sinusoidal driving-voltage signal (i.e., the driving-voltage signal passes through the electrical–acoustic filter), the low and high frequency contents of the driving-voltage are suppressed. The radiated acoustic energy concentrates near the resonance frequency of the source and the radiated acoustic signal (the output signal of the electrical–acoustic filter) has better monochromaticity. This case is suitable for acoustic logging of nondispersive waves, such as the acoustic velocity and amplitude logging of *P*- and *S*- waves.

TABLE III. Parameters of the three kinds of driving-voltage signals.

Gated sinusoidal function		Impulsive function		Boxcar function	
U_0 (V)	t_0	U_0 (V)	δ	U_0 (V)	t_1
200	$6\pi/\omega_1$	22.810×10^{15}	$0.6\omega_1$	400	π/ω_1

In logging, low frequency modes can be used to evaluate fractures and permeability of the formation. Due to limited borehole and logging tool sizes, the lower limit of the minimal resonance frequency of the source only reaches about 11 kHz. Using a general driving-voltage signal to excite the transducer excites low frequency modes poorly. For the conventional acoustic logging tool, we can use the impulsive driving-voltage signal to excite the source. The frequency bandwidth of the acoustic signal radiated by the source is 9.870 kHz. The radiated acoustic energy evenly distributes in the frequency range from 2.185 kHz to 12.055 kHz. Therefore, low frequency modes can be efficiently excited during logging. For this case, the electrical–acoustic conversion efficiency decreases somewhat.

When the transducer is excited by a narrow boxcar driving-voltage signal, the radiated acoustic signal is formed by two wavelets. The first wavelet is caused by appearance of the driving-voltage signal and the second one results from its disappearance. When the two wavelets constructively interfere, the radiated acoustic signal is the greatest. The source creates less heat dissipation during its excitation. For high temperature environments, using the narrow boxcar driving-voltage signal to excite the transducer can improve the reliability and stability of acoustic logging tool.

The source acts like a band-pass electrical–acoustic filter. During its excitation process, the radiated acoustic signal depends on the properties of both the driving-voltage signal and the source.

Because the thin spherical shell transducer radiates acoustic signal evenly in all directions, if the geometrical size of the transducer is considered and $t-r_b/\nu_m$ is used to

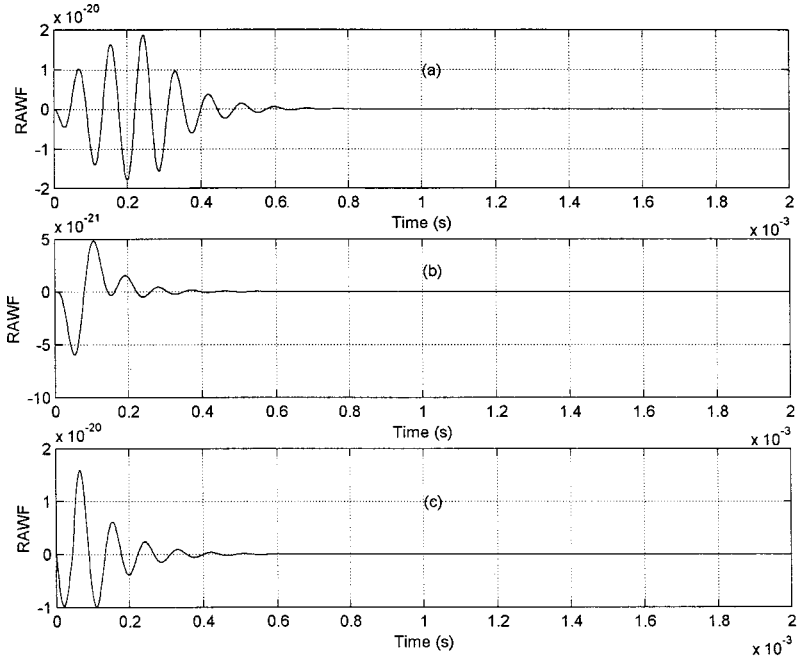


FIG. 5. Time domain wave forms of radiated acoustic signals. RAWF is the amplitude of the radiated acoustic signals. (a) Case of gated sinusoidal driving-voltage signal. (b) Case of impulsive driving-voltage signal. (c) Case of narrow boxcar function driving-voltage signal.

replace t in Eqs. (6)–(8), it can be equivalent to a point source.

C. Calculation of recorded logging signal

From Eq. (2a), it can be known that the receiver is a low-pass filter and has the acoustic–electrical transmission property of the positive original phase. There is a phase difference of π radians between the electrical–acoustic transmission property and the acoustic–electrical transmission property for the transducer. From Eq. (2b), the calculated resonance frequency f_3 of the receiver is equal to 13.100 kHz which is greater than that of the acoustic source. During logging, when the source T_0 is excited by the driving-voltage signal $U_1(t)$ [i.e., when $U_1(t)$ is input the electrical terminals of T_0], $U_1(t)$ is converted into the acoustic signal $X(t,0)$ radiating outwards. $X(t,0)$ passes through the propagation media (borehole fluid and formation around borehole) and then becomes the pressure response signal $P(t,z)$ in the borehole. When $P(t,z)$ is received by the receiver R_0 [i.e., $P(t,z)$ is input to mechanical terminals of the receiver], it is converted into the electrical signal $U_3(t)$ (the recorded logging signal).

Let the radius of the borehole $a=0.1$ m, the separation from source to receiver $L=2.44$ m, and physical parameters of borehole fluid and formation around the borehole as shown in Table IV. The calculated electrical output signals of the receiver in time (ERTW) and frequency (ERAS) domains are shown in Figs. 7–10. When the pressure response signals in borehole are converted into electrical signals by the re-

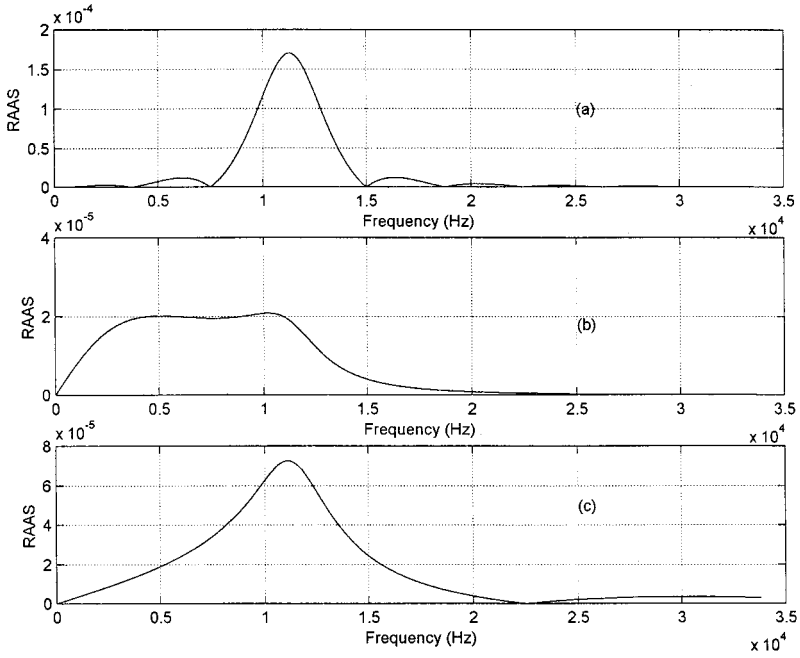


FIG. 6. Amplitude spectra of radiated acoustic signals. RAAS is the amplitude spectrum of the radiated acoustic signals. (a) Case of gated sinusoidal driving-voltage signal. (b) Case of impulsive driving-voltage signal. (c) Case of narrow boxcar function driving-voltage signal.

TABLE IV. Physical parameters of borehole fluid and formation around borehole.

Medium	Density (kg/m^3)	V_p (m/s)	V_s (m/s)
Borehole fluid	1200	1540	
Formation 1	2160	5943	3200
Formation 2	2100	3500	1950

ceiver, its high frequency contents of the signals are filtered and the low frequency contents are suppressed to a certain extent. Compared with the pressure response signals in the borehole, due to the acoustic–electrical filtering effect of the receiver, the wave forms of the recorded logging signals (electrical signal output by the receiver) are smoother. Since the resonance frequency of the receiver is greater than the center frequency of the pressure response in borehole, the peaks of the amplitude spectra of the electric signals output by the receiver become wider. Due to the acoustic–electric transmission property of the positive original phase of the receiver, the first arrivals of the recorded logging signals (head waves) are the negative and are the same as both the radiated acoustic signals and the pressure response signals in the borehole.

For the gated sinusoidal driving-voltage signal, the source radiates more monochromatic acoustic signals. Therefore the maximal value of the recorded logging signals appears in the frequency range between the resonance frequency of the source and that of the receiver.

For the impulsive driving-voltage signal, the acoustic energy radiated by the source evenly distributes over a frequency range from 2.185 kHz to 12.055 kHz. The amplitude spectrum of each recorded logging signal has two extreme values: one appears near to 1.564 kHz and the other is at 11.339 kHz. This fact shows that for the case of a being 0.1 m and L being 2.44 m, the propagation media (borehole fluid and formation around borehole) have two corresponding resonance frequencies. The propagation media have smaller attenuation for the frequency contents near to the above two resonance frequencies and have stronger suppression at other frequencies.

For the narrow boxcar function driving-voltage signal, because its gate width is equal to π/ω_1, the two wavelets of the radiated acoustic signal interfere constructively. The low and high frequencies of the radiated acoustic signal are smaller. Therefore, when the radiated acoustic signals pass the propagation media and the receiver, the low frequency contents of the logging signal output by the receiver is less. Since the radiated acoustic energy mainly distributes near the resonance frequency of the source, the energy distribution of the logging signal mainly concentrates in this frequency.

From the above calculations, it can be concluded that the wave form shape, duration, amplitude, and travel time of the logging signal are determined by the properties of the propagation media, source, receiver, driving-voltage signal, the borehole size, and the separation from source to receiver. Compared with high velocity formation, the moderate velocity formation is of greater low-frequency response and greater acoustic attenuation. When the source is excited by the narrow boxcar function driving-voltage signal with $t_1 = \pi/\omega_1$, the amplitude of the recorded logging signal is the greatest and when the source is excited by the impulsive driving-voltage signal with $0.6\omega_1$, the amplitude of the recorded logging signal is the smallest.

In acoustic logging, the characteristics of recorded logging signal are determined by the type of driving-voltage signal, the electrical–acoustic transmission property of the source, the physical properties of propagation media, the borehole size, the separation from the source to receiver, and the acoustic–electrical transmission property of the receiver. If the driving-voltage signal is more monochromatic between the resonance frequencies of the source and the receiver, the recorded logging signal also has better monochromaticity. If the driving-voltage signal is broad band, when it passes through the acoustic logging transmission network, the bandwidth is reduced. Only the frequencies between the reso-

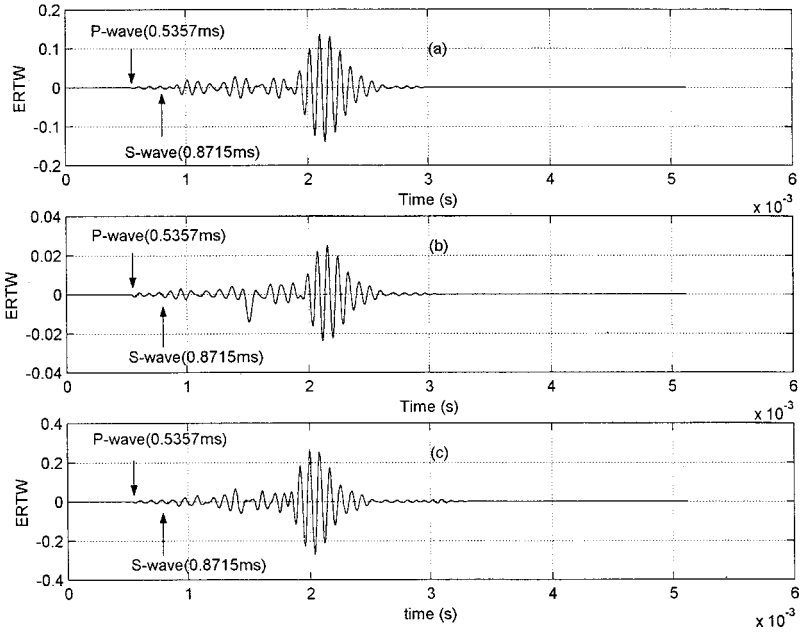


FIG. 7. Wave form of recorded logging signal for formation 1. ERTW is the amplitude of recorded logging signal. (a) Case of gated sinusoidal driving-voltage signal. (b) Case of impulsive driving-voltage signal. (c) Case of narrow boxcar function driving-voltage signal.

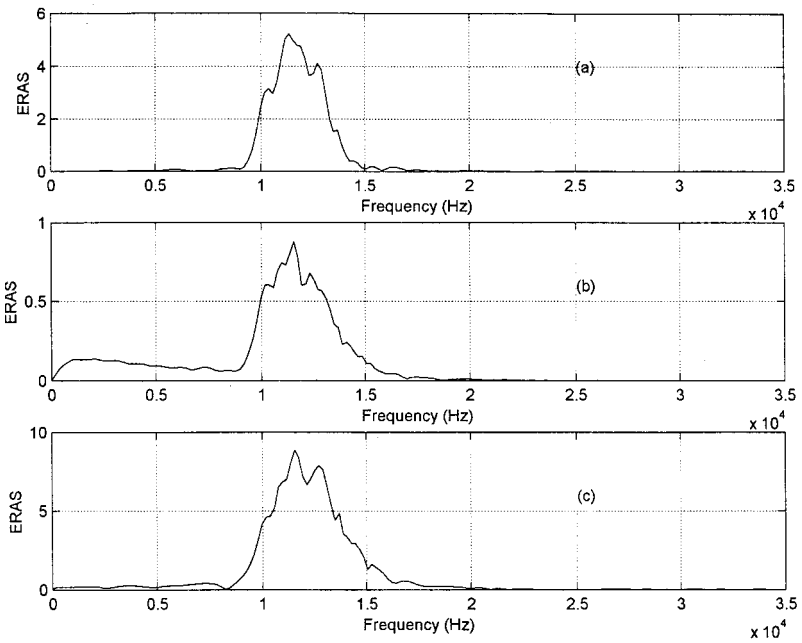


FIG. 8. Amplitude spectrum of recorded logging signal for formation 1. ERAS is the amplitude spectrum of recorded logging signal. (a) Case of gated sinusoidal driving-voltage signal. (b) Case of impulsive driving-voltage signal. (c) Case of narrow boxcar function driving-voltage signal.

nance frequency of the source and that of the receiver can be converted into measured electrical signal efficiently.

IV. CONCLUSIONS AND DISCUSSION

From the above derivation, calculation, simulation, and analysis, some conclusions are as follows.

(1) The source, propagation media and receiver in logging are like electrical–acoustic, acoustic, and acoustic–electrical filters.

(2) The resonance frequency of the receiver is greater than that of the source for the same transducer.

(3) There is a phase difference of π radians between the electrical–acoustic transmission property of the source and the acoustic–electrical transmission property of the receiver. While the radiated acoustic signals propagate in borehole fluid, and the formation around borehole and the pressure response signals in the borehole are converted into the electric signals (i.e., the recorded logging signals), the original phase of their head waves do not change. During the excitation process of the transducer, there is a phase difference of π radians between the original phase of the driving-voltage signal and that of the head wave of the radiated acoustic signal.

(4) There are two resonance frequencies for the propagation media (borehole fluid and formation around borehole). A moderate velocity formation is of greater low frequency response and greater acoustic attenuation compared with a high velocity formation. The acoustic filtering effect and frequency response property of the propagation media are determined by the physical properties of the media and the borehole size.

(5) Compared to the pressure response signals in bore-

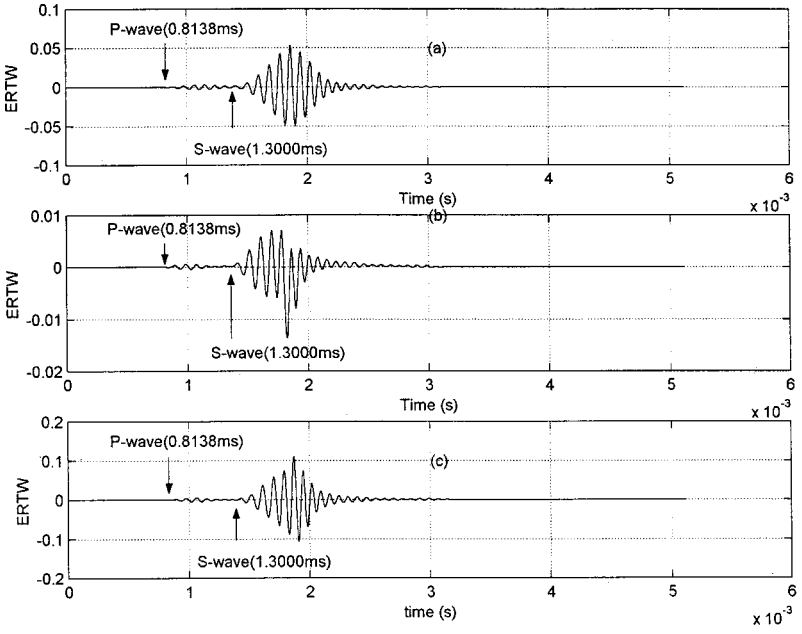


FIG. 9. Wave form of recorded logging signal for formation 2. (a) Case of gated sinusoidal driving-voltage signal. (b) Case of impulsive driving-voltage signal. (c) Case of narrow boxcar function driving-voltage signal.

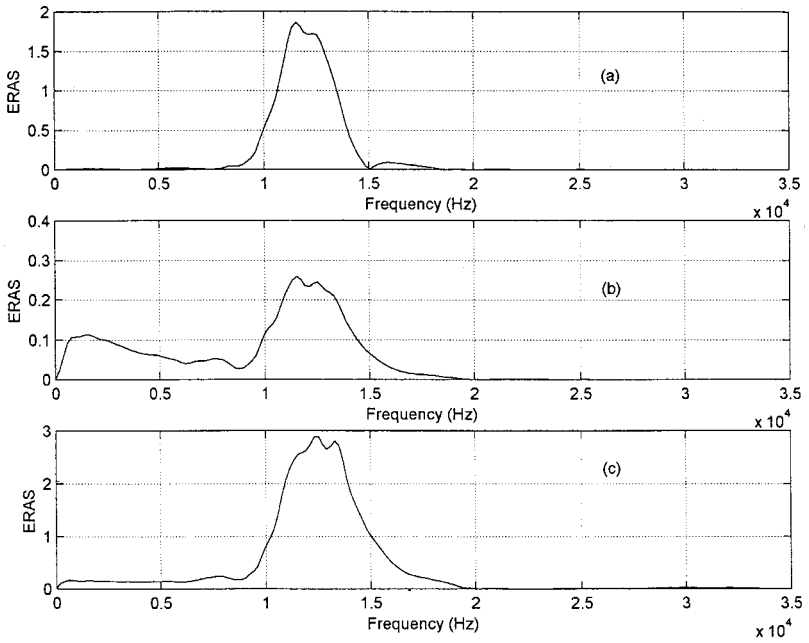


FIG. 10. Amplitude spectrum of recorded logging signal for formation 2. (a) Case of gated sinusoidal driving-voltage signal. (b) Case of impulsive driving-voltage signal. (c) Case of narrow boxcar function driving-voltage signal.

hole, the wave form of the measured logging signal (the electric signals output by the receiver) is smoother and the peaks of its amplitude spectrum are wider.

(6) When the source is excited by a narrow boxcar function driving-voltage signal with $t_1=\pi/\omega_1$, the amplitude of the measured logging signals is the greatest. When the source is excited by the impulsive driving-voltage signal with $0.6\omega_1$, the amplitudes of the pressure response and measured logging signals are the smallest, but they contain more low frequency contents.

(7) The recorded logging signal is the result of the effect of the electrical–acoustic conversion of source, the propagation media, and the acoustic–electrical conversion of receiver on the driving-voltage signal. The recorded logging signal is not only determined by the media, but also by the driving-voltage signal, borehole size and the properties of the source and receiver.

(8) On the basis of the acoustic logging network model, from the recorded logging signals, source property, receiver property, and driving-voltage signal, one can invert for the formation properties more correctly and get more formation information from the recorded logging signals.

ACKNOWLEDGMENT

The authors would like to thank the Institute for Exploration and Development Geosciences, University of Oklahoma for financial support.

[1] K. Aki and P. G. Rechards, *Quantitative Seismology, Theory and Methods* (W. H. Freeman, San Francisco, 1980), pp. 182–183.
[2] L. Tsang and D. Rader, "Numerical evaluation of the transient acoustic waveform due to a point source in a fluid-filled bore hole," Geophysics **44**, 1706–1720 (1979).
[3] R. L. Gibson, Jr. and C. Peng, "Low- and high-frequency radiation from seismic sources in cased boreholes," Geophysics **59**, 1780–1785 (1994).
[4] C. H. Cheng and M. Nafi Toksöz, "Elastic wave propagation in a fluid-filled borehole and synthetic acoustic logs," Geophysics **46**, 1042–1053 (1981).
[5] D. A. Berlincourt, D. R. Curran, and H. Jaffe, "Piezoelectric and piezomagnetic materials and their function in transducers," in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1964), Vol. 1A, pp. 223–224.
[6] J. C. Piquette, "Method for transducer suppression. I: Theory," J. Acoust. Soc. Am. **92**, 1203–1213 (1992).
[7] J. C. Piquette, "Method for transducer suppression. II: Experiment," J. Acoust. Soc. Am. **92**, 1214–1221 (1992).
[8] J. C. Piquette, "Applications of the method for transducer suppression to various transducer types," J. Acoust. Soc. Am. **94**, 646–651 (1993).
[9] J. C. Piquette and S. E. Forsythe, "Transducer transient suppression: Generalized methods of analysis," J. Acoust. Soc. Am. **100**, 1577–1583 (1996).
[10] A. C. Holly, "A method for the generation of broadband acoustic transmission," J. Acoust. Soc. Am. **75**, 973–976 (1984).
[11] L. Fa, F. Lin, and W. Chen, "Source function derivation and property analysis of ceramic ring transducer for petroleum exploration," Acta Geophys. Sinica **39**, Suppl. 387–399 (1996).
[12] L. Fa, J. P. Castagna, and J. M. Hovem, "Derivation and simulation of source function for acoustic logging," 1999 IEEE International Ultrasonics Symposium Proceedings, Vol. 1, pp. 707–710.
[13] L. Fa, L. An, and P. Li, "Theoretical relationship electric-driving signal and radiated acoustic wave for thin shell transducer," Well Logging Technol. **22**, 1004–1338 (1998).

Acoustic tracking of a freely drifting sonobuoy field

Stan E. Dosso
School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

Nicole E. B. Collison
Defence Research Establishment Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

(Received 23 July 2001; revised 17 December 2001; accepted 4 February 2002)

This paper develops an acoustic inversion algorithm to track a field of freely drifting sonobuoys using travel-time measurements from a series of nonsimultaneous impulsive sources deployed around the field. In this scenario, the time interval between sources can be sufficiently long that significant independent movement of the individual sonobuoys occurs. In addition, the source transmission instants are unknown, and the source positions and initial sonobuoy positions are known only approximately. The formulation developed here solves for the track of each sonobuoy (parametrized by the sonobuoy positions at the time of each source transmission), allowing arbitrary, independent sonobuoy motion between transmissions, as well as for the source positions and transmission instants. This leads to a strongly underdetermined inverse problem. However, regularized inversion provides meaningful solutions by incorporating *a priori* information consisting of prior estimates (with uncertainties) for the source positions and initial sonobuoy positions, and a physical model for preferred sonobuoy motion. Several models for sonobuoy motion are evaluated, with the best results obtained by minimizing the second spatial derivative of the tracks to obtain the minimum-curvature or smoothest track, subject to fitting the acoustic data to a statistically appropriate level. [DOI: 10.1121/1.1466867]

PACS numbers: 43.60.Pt, 43.30.Xm [DLB]

I. INTRODUCTION

This paper develops an acoustic inversion algorithm to track a field of freely drifting sonobuoys using travel-time measurements from a series of nonsimultaneous impulsive sources deployed around the field. In general, the sonobuoys drift from their initial positions due to the effects of winds and currents; in littoral waters, variable local conditions can result in significant independent motion of the buoys. Hence, to make use of the sonobuoy field for any length of time requires a method for localizing and tracking the individual sensors.[1–4] Ideally, the impulsive sources themselves can be used to provide acoustic data for sonobuoy tracking. The difficulties with this approach are that the positions of the sources are not well known, and the time interval between successive source deployments can be long enough to allow significant movement of the sonobuoys. Traditional acoustic approaches typically simplify the problem by assuming that the source positions are known precisely and that the sonobuoys remain stationary during at least four successive source deployments (the minimum number required for localization using relative travel-time measurements). However, these assumptions may not be valid and can preclude meaningful solutions. A nonacoustic approach is to track one sonobuoy (e.g., using on-board GPS) and assume that all sonobuoys in the field drift in an identical manner. However, this can also be a poor assumption, as the sonobuoys are often observed to move independently due to local variation in wind and currents.

The goal of this paper is to develop an acoustic inversion algorithm to determine individual sonobuoy tracks as well as source positions and transmission instants. The sonobuoy tracks are parametrized by the buoy positions at the time of each source deployment, allowing arbitrary, independent sonobuoy motion between source deployments. This formulation leads to a strongly underdetermined inverse problem. For example, a scenario consisting of a field of N_1 sonobuoys and N_2 source transmissions yields a data set of N_1N_2 direct-path arrival times, but involves $3N_1N_2+4N_2$ unknown parameters ($3N_1N_2$ parameters to represent the 3D position of each sonobuoy at the time of each source transmission, $3N_2$ parameters to represent the source positions, and N_2 parameters to represent the unknown source transmission instants). The number of data can be increased by including the travel times of bottom-reflected signals; however, as these signals arrive at the sonobuoys at the same horizontal angles as the direct arrivals, they typically provide little independent information.

In principle, the difficulty with underdetermined inverse problems is not that no solution exists which fits the data (typically the case for overdetermined problems), but rather that an infinite number of solutions exist which fit the data. Further, these solutions can be highly diverse, and many may not be physically realistic. In this case, the goal is to formulate a unique inversion for a physically meaningful solution. The method of regularization[5–9] provides a powerful approach to such problems by incorporating physical *a priori* information regarding the solution into the inverse problem. In particular, in the algorithm developed here, the source positions, initial sonobuoy positions, and sonobuoy depths are treated as unknown parameters, but are regularized by including *a priori* estimates, with uncertainties, such as would typically be available from the deployment procedure.

The sonobuoy tracks are regularized by specifying a preferred model for sonobuoy motion, with several different models considered here. The simplest model for sonobuoy motion is no motion, i.e., the sonobuoys remain approximately fixed at their initial positions. A more physically meaningful prior model is that the sonobuoys move in response to wind and currents along smooth tracks, i.e., with minimal curvature or changes in direction. However, if motion along curved paths is expected, a prior model consisting of quadratically curved tracks (which includes linear tracks) can be applied. It is important to note that in the regularized approach, the *a priori* information is applied subject to fitting the measured data to a statistically appropriate level. Hence, the prior position estimates and the preferred model for sonobuoy motion will be satisfied only to the extent that they agree with the travel-time measurements; the solution will include any differences from the *a priori* information that are resolved by the data.

Regularized inversion has been applied previously to localize the sensors of vertical and horizontal line arrays using travel-time measurements from a series of sources positioned around the array, but with the source positions only approximately known.[10–12] These problems, although formally overdetermined (more data than unknowns), are ill-conditioned due to the source position uncertainty. More recently, regularized inversion was applied to determine the depth and offset of individual sensors of a ship-towed marine seismic array using direct and bottom-reflected acoustic arrivals from a towed seismic source.[13] This problem was weakly underdetermined since the direct and/or reflected arrivals were not present at some sensors due to the shadowing effects of water-column refraction or reflection from arbitrary bathymetry. In this case, regularized inversion for the smoothest array shape provided the requisite *a priori* information to localize the shadowed sensors. The new inversion developed here for physically constrained sonobuoy tracks allowing independent movement between successive source transmissions would appear to be the first application of regularized inversion to such a strongly underdetermined localization problem.

The remainder of this paper is organized as follows. Section II A presents a brief overview of the necessary results from inverse theory, and Sec. II B describes the implementation of the regularized tracking algorithm. Section III A illustrates the tracking algorithm with a realistic synthetic example. Section III B compares the results for the various regularizations (i.e., models for sonobuoy motion) and choices of data sets (with and without bottom-reflected arrivals). Section III C examines the sensitivity to errors in the data and in the prior estimates, and the effect of the geometric configuration of source positions. The work is summarized in Sec. IV. Finally, the Appendix describes the ray-theory implementation used in the tracking algorithm.

II. SONOBUOY TRACKING ALGORITHM

A. Inverse theory

This section summarizes the inverse theory that forms the basis for the sonobuoy tracking algorithm; more general treatments of the subject can be found in Refs. 5–9. The implementation of the inversion algorithm is described in Sec. II B and in the Appendix. The acoustic data represent measured arrival times of direct and possibly bottom-reflected acoustic signals measured at the sonobuoys due to multiple sources. The source transmission instants are not known in the present application, but are considered parameters to be determined in the inversion. Hence, the arrival-time data represent relative travel time information (relative to any arbitrary clock reference), i.e., it is the time difference between arrivals that provides information. The acoustic data $\mathbf{t}$ can be written in general vector form as

$$\mathbf{t}=\mathbf{t}(\mathbf{m})+\mathbf{n}. \tag{1}$$

In (1), the model $\mathbf{m}$ of unknown parameters is taken to consist of the 3D position (x, y, and z) of each sonobuoy at the time of each source transmission, as well as the position and transmission instant of each source. The forward mapping $\mathbf{t}(\mathbf{m})$ represents the acoustic travel times along direct or bottom-reflected ray paths between source and receivers (described in the Appendix). Finally, $\mathbf{n}$ represents the data errors (noise). The inverse problem of determining $\mathbf{m}$ from $\mathbf{t}$ is functionally nonlinear; however, a local linearization is obtained by expanding $\mathbf{t}(\mathbf{m})=\mathbf{t}(\mathbf{m_0}+\delta\mathbf{m})$ in a Taylor series to first order about an arbitrary starting model $\mathbf{m_0}$ to yield

$$\mathbf{t}=\mathbf{t}(\mathbf{m_0})+\mathbf{J}\delta\mathbf{m}, \tag{2}$$

where $\delta\mathbf{m}$ represents an unknown model perturbation and $\mathbf{J}$ is the Jacobian matrix consisting of the partial derivatives of the data functionals with respect to the model parameters, evaluated at the starting model

$$J_{ij}=\partial t_i(\mathbf{m_0})/\partial m_j \tag{3}$$

(partial differentiation of the ray travel-time function is considered in the Appendix). Substituting $\delta\mathbf{m}=\mathbf{m}-\mathbf{m_0}$, the expansion can be written

$$\mathbf{Jm}=\mathbf{t}-\mathbf{t}(\mathbf{m_0})+\mathbf{Jm_0}\equiv\mathbf{d}. \tag{4}$$

Equation (4) defines a linear inverse problem for $\mathbf{m}$ in terms of known quantities (defined as modified data $\mathbf{d}$), which can be solved using methods of linear inverse theory (described below). Since nonlinear terms are neglected in (4), the equations are not exact and the solution $\mathbf{m}$ may not adequately reproduce the measured data via the nonlinear forward problem. In this case, the starting model is updated, $\mathbf{m_0}\leftarrow\mathbf{m}$, and the inversion repeated iteratively until an acceptable solution is obtained. Note that since (4) is formulated in terms of the model $\mathbf{m}$, rather than the perturbation $\delta\mathbf{m}$, *a priori* information regarding the model can be applied directly to the inversion, as required for the present application.

If the linear inverse problem given by Eq. (4) is overdetermined, then under the standard assumption that the error n_i on datum t_i is due to an independent, Gaussian-distributed random process with zero mean and standard deviation σ_i, the maximum-likelihood solution is obtained by minimizing the χ^2 data misfit

$$\chi^2=|\mathbf{G}(\mathbf{Jm}-\mathbf{d})|^2, \tag{5}$$

where $\mathbf{G}=\mathrm{diag}[1/\sigma_i]$, to obtain

$$\mathbf{m}=[\mathbf{J}^T\mathbf{G}^T\mathbf{G}\mathbf{J}]^{-1}\mathbf{J}^T\mathbf{G}^T\mathbf{G}\mathbf{d}. \tag{6}$$

However, this least-squares approach is not applicable to underdetermined inversions, such as the sonobuoy-tracking problem at hand. In such cases, an infinite number of solutions exist that fit the data exactly: the numerical manifestation of this indeterminacy is that the matrix in square brackets in (6) is singular.

A particularly useful approach to underdetermined inversions is given by the method of regularization, which explicitly includes physical *a priori* information regarding the solution in the inversion to define a well-posed inverse problem. This is accomplished by minimizing an objective function ϕ that combines the χ^2 data misfit with a regularizing term that imposes the *a priori* expectation that the model $\mathbf{m}$ resembles a prior estimate $\hat{\mathbf{m}}$

$$\phi=|\mathbf{G}(\mathbf{J}\mathbf{m}-\mathbf{d})|^2+\mu|\mathbf{H}(\mathbf{m}-\hat{\mathbf{m}})|^2, \tag{7}$$

where $\mathbf{H}$ is the regularization matrix (described below), and μ is a trade-off parameter (Lagrange multiplier) controlling the relative importance assigned to the data misfit and the *a priori* information in the minimization. Minimizing ϕ with respect to $\mathbf{m}$, the regularized solution is

$$\mathbf{m}=\hat{\mathbf{m}}+[\mathbf{J}^T\mathbf{G}^T\mathbf{G}\mathbf{J}+\mu\mathbf{H}^T\mathbf{H}]^{-1}[\mathbf{J}^T\mathbf{G}^T\mathbf{G}\mathbf{d}-\mathbf{J}\hat{\mathbf{m}}]. \tag{8}$$

Note that for suitable choices of μ and $\mathbf{H}$, the presence of the term $\mu\mathbf{H}^T\mathbf{H}$ within the square brackets in (8) ensures that the matrix inversion is well-conditioned.

The regularization matrix $\mathbf{H}$ in (7) and (8) controls the form of the *a priori* information applied in the inversion. For instance, if prior model parameter estimates $\hat{\mathbf{m}}$ are available, the appropriate regularization is given by a diagonal matrix

$$\mathbf{H}=\mathrm{diag}[1/\xi_j], \tag{9}$$

where ξ_j represents the uncertainty for the jth parameter estimate $\hat{m}_j$. An alternative form of regularization is to apply *a priori* information to derivatives of the model parameters. For instance, if the *a priori* expectation is that the parameters remain approximately constant, then an appropriate choice is $\hat{\mathbf{m}}=\mathbf{0}$ and

$$\mathbf{H}=\begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{10}$$

Each row of $\mathbf{H}$ in (10) represents a discrete approximation to the first derivative operator. Hence, minimizing $|\mathbf{H}\mathbf{m}|^2$ minimizes the l_2 norm of the model derivative. The result is the model that varies minimally from a constant value, often referred to as the flattest model. Alternatively, if the *a priori* expectation is that the parameters vary in a smooth manner, then an appropriate choice is $\hat{\mathbf{m}}=\mathbf{0}$ and

$$\mathbf{H}=\begin{bmatrix} -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{11}$$

Each row of $\mathbf{H}$ in (11) represents a discrete approximation to the second derivative operator; hence, $|\mathbf{H}\mathbf{m}|^2$ provides a measure of the total curvature of the model, often referred to as the model roughness. Applying this regularization produces the minimum-curvature solution, referred to as the smoothest model. Finally, if the *a priori* expectation is that the parameters approximate a quadratic function (which includes linear tracks as a subcase) then an appropriate choice is $\hat{\mathbf{m}}=\mathbf{0}$ and an approximation to the third derivative operator

$$\mathbf{H}=\begin{bmatrix} -1 & 3 & -3 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 3 & -3 & 1 & \cdots & 0 & 0 & 0 & 0 & 0 \\ \vdots & & & & & \ddots & & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 3 & -3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 3 & -3 & 1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{12}$$

For any of the above regularizations, the trade-off parameter μ in (8) is chosen so that the χ^2 data misfit achieves its expected value of $\langle\chi^2\rangle=N$ for N data, thereby applying the *a priori* information subject to ensuring that the data are fit to a statistically appropriate level.

Note that the regularization matrices in (10), (11), and (12) involve one, two, and three rows of zeros, respectively, which represent constants of integration that are not constrained by the regularization. For instance, the regularization matrix in (10) constrains the first derivative of the

model; to translate this to a constraint on the model itself requires an integration constant representing the model value at some fixed point. Since this information is not provided in the regularization, it must be determined by the data. Likewise, constraining the third derivative via (12) requires that three integration constants (values of the model and its first and second derivatives) be determined by the data. Hence, the *a priori* information content of the regularization decreases with the order of the derivative constrained. It is also clear from (8) that the rows of zeroes in $\mathbf{H}$ (and therefore in $\mathbf{H}^T\mathbf{H}$) degrade the conditioning of the matrix inversion and can lead to an ill-posed (unstable) problem.

The sonobuoy tracking problem formulated here applies *a priori* information in terms of prior parameter estimates and constraints on parameter derivatives. In particular, prior parameter estimates (with appropriate uncertainties) are applied to the source positions, the initial sonobuoy positions, and the sonobuoy depths. In addition, one of the above derivative regularizations is applied to the horizontal sonobuoy tracks, providing a preferred model for sonobuoy motion. Applying the first-derivative regularization using (10) assumes that the sonobuoy positions remain essentially constant near their initial estimates. The second-derivative regularization (11) assumes that the sonobuoys tend to move along smooth tracks with minimal changes in direction. The third-derivative regularization (12) assumes that the sonobuoys move along curved (quadratic) tracks. Each of these three models for sonobuoy motion is evaluated in Sec. III B. Note that in the regularized approach, the *a priori* information is applied subject to fitting the acoustic travel times; hence, the prior assumptions (constant positions, or smooth or quadratic tracks) will not be exactly satisfied if a different track structure can be resolved by the data.

To apply two different types of *a priori* information simultaneously, an augmented objective function can be formed which includes two regularization terms

$$\phi = |\mathbf{G}(\mathbf{Jm}-\mathbf{d})|^2 + \mu_1|\mathbf{H}_1(\mathbf{m}-\hat{\mathbf{m}}_1)|^2 + \mu_2|\mathbf{H}_2(\mathbf{m}-\hat{\mathbf{m}}_2)|^2. \tag{13}$$

In (13), the first regularization term is taken to represent the *a priori* parameter estimates for the source positions, the initial sonobuoy positions, and sonobuoy depths. Hence, $\hat{\mathbf{m}}_1$ consists of the prior estimates for these parameters, with zeros for the remaining parameters. The regularization matrix $\mathbf{H}_1$ is of the form of (9) with diagonal elements consisting of the reciprocal of the estimated uncertainty for parameters with prior estimates, and zeros for the remaining parameters. The second regularization term is taken to represent one of the models for sonobuoy motion described above. Hence, $\hat{\mathbf{m}}_2$ is taken to be zero, and $\mathbf{H}_2$ is given by one of (10), (11), or (12) for the horizontal sonobuoy position parameters. Rows of zeros are included in $\mathbf{H}_2$ at appropriate locations to separate the measures of curvature in x and y. In this case, minimizing (13) leads to

$$\mathbf{m} = \hat{\mathbf{m}}_1 + [\mathbf{J}^T\mathbf{G}^T\mathbf{GJ} + \mu_1\mathbf{H}_1^T\mathbf{H}_1 + \mu_2\mathbf{H}_2^T\mathbf{H}_2]^{-1} \times [\mathbf{J}^T\mathbf{G}^T\mathbf{Gd} - \mathbf{J}\hat{\mathbf{m}}_1]. \tag{14}$$

Lagrange multipliers μ_1 and μ_2 are chosen to provide a statistically appropriate fit to the data and prior information, as described in the following section.

B. Algorithm implementation

The implementation of the sonobuoy tracking algorithm consists of an iterative application of the dual regularized solution (14), initiated from a starting model coinciding with the prior parameter estimates. Numerical convergence of the algorithm is based on: (i) obtaining a statistically meaningful misfit to the measured data of $\chi^2 = N$, and (ii) obtaining a stable solution such that the root-mean-square (rms) change in the sensor positions between iterations is negligible compared to the accuracy required in the inversion (a value of 0.1 m was adopted here). Regarding the first criterion, note that although (14) is derived based on the χ^2 misfit for the linear inverse problem (4) that approximates the nonlinear problem (1) at each iteration, the convergence of the inversion algorithm must be judged in terms of the nonlinear misfit

$$\chi^2 = |\mathbf{G}(\mathbf{t}(\mathbf{m}) - \mathbf{t})|^2. \tag{15}$$

An equivalent and sometimes more convenient measure is the rms data misfit

$$X = [\chi^2/N]^{1/2}, \tag{16}$$

with an expected value $\langle X \rangle = 1$.

The most subtle aspect of implementing the inversion involves assigning values to the two trade-off parameters, μ_1 and μ_2, which control the balance between the data misfit and the two forms of *a priori* information. An effective procedure[10,13] is to set

$$\mu_2 = \beta\mu_1 \tag{17}$$

for a fixed value of β, and determine the value of μ_1 at each iteration which yields the desired rms data misfit X. The parameter β provides a relative weighting between fitting the prior estimates (emphasized with small β) and minimizing the appropriate model derivative (emphasized with large β). The final model obtained from this procedure can then be examined to ascertain whether the value of β was appropriate based on the rms parameter misfit associated with the *a priori* estimates, defined by

$$\hat{X} = [|\mathbf{H}_1(\mathbf{m}-\hat{\mathbf{m}}_1)|^2/\hat{M}]^{1/2}, \tag{18}$$

where $\hat{M}$ is the number of model parameters with *a priori* estimates. To fit the prior estimates to a statistically meaningful level requires $\hat{X} \approx 1$. If $\hat{X} \ll 1$, then the prior information is fit more closely than is reasonable statistically, and a larger value of β is required to reduce the relative weighting of the prior information. Conversely, if $\hat{X} \gg 1$, the prior information is not fit closely enough, and a smaller value of β is required to increase the weighting of the prior information. The inversion can be repeated with a new value of β until $\hat{X} \approx 1$ is achieved. In practice, determining an appropriate value for β is a straightforward procedure, typically requiring two or three trial inversions for a particular problem. The value of β usually remains constant when inverting succes-

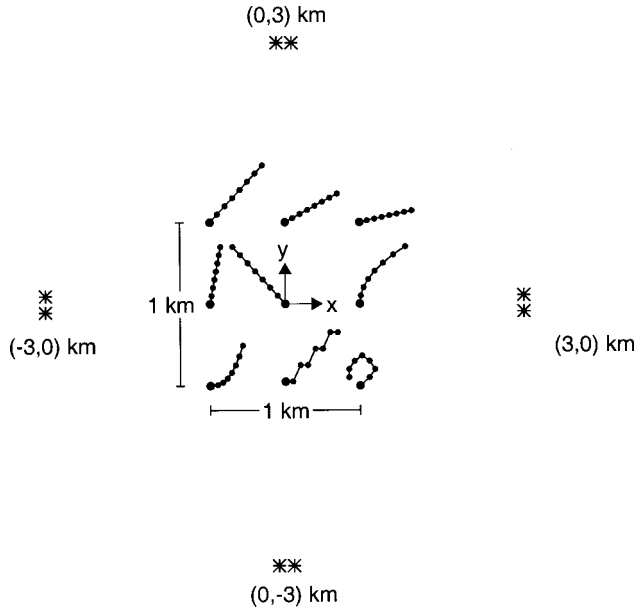


FIG. 1. Schematic plan-view diagram of the geometry of the synthetic sonobuoy-tracking example. Connected dots represent sonobuoy tracks, with the larger dots indicating the initial sonobuoy position. Asterisks indicate acoustic source positions.

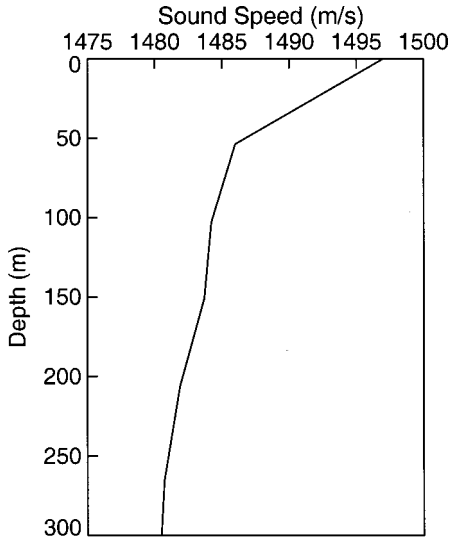


FIG. 2. Ocean sound-speed profile for the sonobuoy-tracking example.

sive data sets with similar uncertainties; hence, the algorithm is appropriate for batch processing.

The above procedure reduces the problem of determining two trade-off parameters to a sequence of 1D searches for the parameter μ_1. The parameter μ_1 is chosen so that X is reduced by a controlled amount (e.g., a factor of 5) at each iteration until $X=1$ is achieved. Controlling the reduction in X limits the change in the model at each iteration, which helps ensure that the linearization is valid and stabilizes the convergence. In practice, it is straightforward to determine the value of μ_1 which produces the desired X at a given iteration since X increases monotonically with μ_1. At early iterations an approximate value for μ_1 is sufficient, and a bisection algorithm is employed. Near convergence, Newton's method is used for greater precision.

III. EXAMPLES

A. Canonical test case

This section illustrates the acoustic inversion for sonobuoy tracking with a realistic synthetic example. Figure 1 shows a schematic plan-view diagram of the scenario, which involves a field of nine freely drifting sonobuoys with eight acoustic sources deployed around the perimeter of the field. The track of each sonobuoy is indicated by a series of connected dots, with each dot representing the sonobuoy position at the time of transmission of each of the sources. The larger dots indicate the initial position of each sonobuoy (i.e., the sonobuoy position at the transmission time of the first source). The initial positions are separated by 500 m in x and y, and define a 3×3 grid centered on the coordinate origin. The depth of all sonobuoy sensors is taken to be 100 m. Sonobuoys 1–5 (ordered from left-to-right, top-to-bottom) follow linear tracks at differing horizontal angles and drift rates. Sonobuoys 6 and 7 follow curved tracks. The track of sonobuoy 8 describes a zigzag pattern, and sonobuoy 9 follows an approximately circular path as if caught in a current gyre. The variety of sonobuoy tracks in this test case would likely not be observed in a single sonobuoy field of these dimensions; however, it was chosen to illustrate the ability of the inversion to resolve diverse, independent sonobuoy motion. The source positions are indicated by asterisks in Fig. 1. There are a total of eight source transmissions from four source positions arranged around the sonobuoy field. The first four sources are positioned at $(x,y)=(0,3)$, $(3,0)$, $(0,-3)$, and $(-3,0)$ km, respectively; these positions are then repeated for the last four sources (representative of sources deployed by a circling aircraft). The depth of the sources is taken to be 180 m.

The ocean environment for this test case involves a typical shallow-water sound-speed profile, as shown in Fig. 2. The synthetic acoustic data consist of relative travel times for direct arrivals from sources to receivers, computed using the ray-tracing algorithm outlined in the Appendix. Random errors drawn from a zero-mean, Gaussian distribution with a standard deviation of 1 ms were added to the data. The prior information included in the inversion algorithm was comprised of estimates of the source positions, initial sonobuoy positions, and sonobuoy depths at all positions along the track. The source-position estimates included random Gaussian errors of standard deviations 50 m in x and y and 5 m in z. The initial sonobuoy-position estimates included Gaussian errors with standard deviations of 100 m in x and y, and all sonobuoy depth estimates included Gaussian errors of 5 m.

The regularized inversion described in Sec. II was applied to the data set and prior information described above. The smoothest (minimum-curvature) sonobuoy tracks that fit the acoustic data were determined by applying the regularization given by Eq. (11), since this would seem to be the most physically meaningful model for sonobuoy motion (other choices are consider later in this paper). The starting model consisted of the *a priori* estimates for the source positions and initial sonobuoy positions. Since no prior knowledge is assumed about drift directions, the starting values for

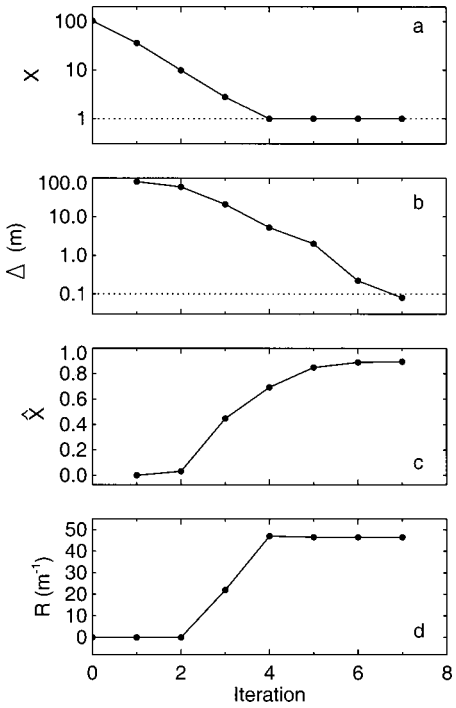


FIG. 3. Convergence properties of the regularized inversion algorithm for the sonobuoy-tracking example. See the text for detailed explanation.

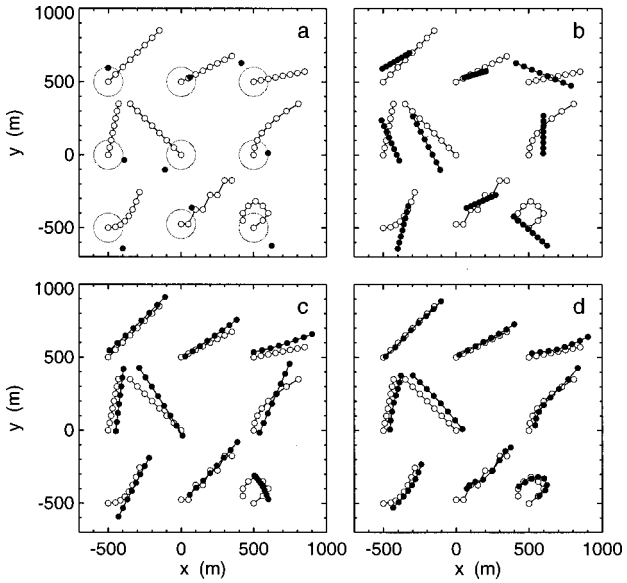


FIG. 4. Sonobuoy tracks constructed at various stages of the inversion algorithm. True tracks are represented by open circles. Constructed tracks are represented by filled circles. Iterations 0 (starting model), 1, 2, and 7 (final model) are shown in (a)–(d), respectively. The gray circles in (a) indicate the uncertainty (standard deviation) in the prior estimates for the initial sonobuoy positions.

the subsequent sonobuoy positions along each track were also taken to coincide with the initial positions (i.e., the starting model assumed all sonobuoys remained fixed at their initial positions). The convergence of the inversion algorithm for this case is illustrated in Fig. 3 in terms of the rms data misfit X, the rms change in sonobuoy positions between iterations Δ, the rms misfit to the prior parameter estimates $\hat{X}$, and the array roughness $R=|\mathbf{H}_2\mathbf{m}|^2$, where $\mathbf{H}_2$ is given by (11). The starting model (iteration 0) has a large data misfit of $X\approx 100$, as shown in Fig. 3(a). As the inversion progresses, X decreases by a factor of approximately 5 at each iteration, until the desired value of unity is obtained at iteration 4. The data misfit X then remains fixed at this value until the model change Δ decreases below the threshold of 0.1 m for convergence at iteration 7 [Fig. 3(b)]. The prior misfit $\hat{X}$ increases from $\hat{X}=0$ at the starting model to a value of $\hat{X}\approx 0.9$ at convergence [Fig. 3(c)], which approximates the expected value of unity, as described in Sec. II. Similarly, the track roughness R increases from an initial zero value as structure is built into the tracks to fit the data, reaching a stable value by iteration 4.

The sonobuoy tracks at selected iterations of the inversion process are shown in Fig. 4. The open circles depict the true sonobuoy positions along the drift track. The filled circles indicate the estimated sonobuoy positions. Figure 4(a) shows the starting model (iteration 0). As previously mentioned, the starting model for all sonobuoy positions along each track consisted of the estimated initial sonobuoy position; hence, each filled circle in Fig. 4(a) represents the entire track. The gray circles in Fig. 4(a) indicate the uncertainty (standard deviation) about the true initial sonobuoy positions (note that since the errors are Gaussian distributed, the initial position estimates can and do lie outside the gray circles). Figure 4(b) shows the estimated sonobuoy positions after the first iteration. The rudimentary tracks at this stage clearly indicate a sensitivity to sonobuoy motion in approximately the correct directions. By the second iteration [Fig. 4(c)], the linear tracks are fairly well determined, but the curved tracks are not. Figure 4(d) shows the final model at convergence (iteration 7). Even though there was no prior information about the specific movement of individual sonobuoys, the final solution is a good approximation to the true sonobuoy tracks in all cases, although the linear tracks are typically fit more closely than the curved or zigzag tracks. A virtually identical result was obtained by initiating the algorithm from a wide range of starting models, indicating the inversion is not sensitive to the choice of initialization (i.e., the parameter space does not seem to contain local minima as the problem is not strongly nonlinear). The complete inversion required less than 10-s computation time on a 450-MHz PC running IDL (Interactive Data Language). Hence, in practice, once a suitable number of source transmissions (four or more) have been recorded, the positioning computations can be carried out in essentially real time, and updated as further transmissions are recorded.

The final solution for the sonobuoy tracks, shown in Fig. 4(d), was calculated for a particular realization of random errors on the data and prior parameter estimates. The results of the inversion vary somewhat depending on the particular error realizations. To quantify reliably the uncertainty in sonobuoy positions, a Monte Carlo appraisal was carried out based on performing 50 independent inversions, each with different random errors on the data and prior parameter estimates. The results are characterized in terms of error ellipses defined by the standard deviations of the sonobuoy x and y position solutions about the true positions, as shown in Fig. 5. The uncertainty estimate for each sonobuoy position is

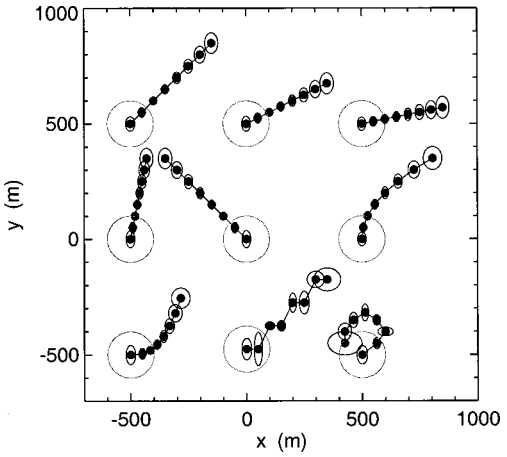


FIG. 5. Uncertainty ellipses computed via Monte Carlo appraisal analysis for smoothest sonobuoy tracks using direct-arrival data. Data errors are 1 ms, prior estimate errors are 50 m for source positions, 100 m for initial sonobuoy positions (gray circles), and 5 m for source and sensor depths.

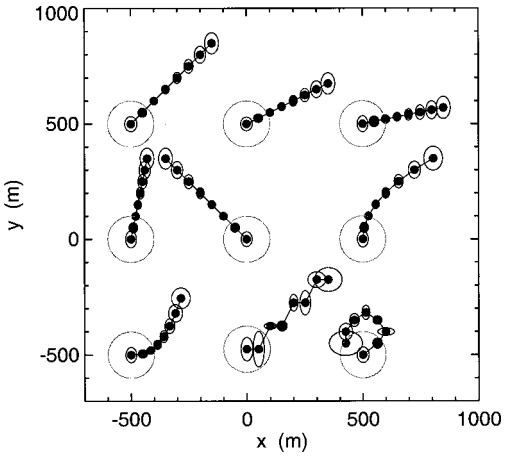


FIG. 6. Results of Monte Carlo appraisal analysis for smoothest sonobuoy tracks using direct- and reflected-arrival data. Data errors are 1 ms, prior estimate errors are 50 m for source positions, 100 m for initial sonobuoy positions (gray circles), and 5 m for source and sensor depths.

taken to be the semimajor axis of the corresponding error ellipse (i.e., the larger of the x or y error). The average uncertainty over all sonobuoy positions in Fig. 5 is 32 m, and the maximum uncertainty is 75 m. These uncertainties are substantially smaller than the uncertainty of the prior estimate of the initial sonobuoy positions (100 m). The sonobuoy position uncertainties are typically larger at the start and end of the tracks and smaller at intermediate positions, likely due to the end effects of the regularization for smoothest sonobuoy tracks given by Eq. (11). The uncertainties computed for the sonobuoy depths and for the source positions and depths were found simply to reflect the prior estimate uncertainties (i.e., these parameters are unconstrained by the data), and are not considered further here.

In the remainder of this paper, a number of variations on the basic test case presented in this section (referred to as the canonical test case) are considered to investigate different aspects of the inversion. In each case, the results are presented in terms of sonobuoy track uncertainties as computed via Monte Carlo appraisal analysis, with average and maximum position errors given in Table I.

TABLE I. Comparison of average and maximum sonobuoy position errors for the canonical example (described in Sec. III A), and other cases that vary different aspects of this example. Position errors represent the larger of the x and y errors; averages and maxima are taken over all sonobuoy positions for a given case.

		Position error (m)	
Case	Figure	Average	Max
Canonical test case	5	32	75
Direct and reflected data	6	31	77
First-derivative regularization	7(a)	97	220
Third-derivative regularization	7(b)	255	849
10× data errors (10 ms)	8(a)	46	92
1/10 data error (0.1 ms)	8(b)	32	79
2× prior estimate errors	9(a)	55	92
1/2 prior estimate errors	9(b)	23	64
Circular source configuration (2 km)	10(a)	36	75
Semicircular source configuration (2 km)	10(b)	67	221

B. Choice of data and regularization

The tracking results for the canonical test case, shown in Fig. 5, were computed by inverting a data set consisting of the relative travel times of direct acoustic arrivals. To investigate the advantage of augmenting the data set by including travel-time measurements for bottom-reflected arrivals, a Monte Carlo appraisal analysis of this case was carried out. The analysis included data corresponding to reflections from a flat seabed in addition to the direct arrivals, but was otherwise identical to the appraisal of the canonical test case. The results of the appraisal for the augmented data set, given in Fig. 6, are very similar to the those of the direct-arrival appraisal (Fig. 5). In addition, the average and maximum positioning errors of the two cases, given in Table I, are almost identical. The bottom-reflected arrivals follow identical horizontal paths as the direct arrivals, and hence provide little additional constraint on sonobuoy positions in x and y. However, bottom-reflected signals arrive at different vertical angles, and hence can provide additional z information. In the present application, the sonobuoy depths are assumed to be relatively well known *a priori* (5-m uncertainty); hence, including bottom-reflected data does not significantly improve the results. In other applications where the receiver depths are not well known, including bottom-reflected data can be important.[13]

The inversion results shown to this point were computed using the second-derivative regularization, given by Eq. (11), to produce the smoothest or minimum-curvature sonobuoy tracks. This would seem to a physically realistic *a priori* model for sonobuoy motion. However, to investigate the other possible regularizations, Monte Carlo appraisal was also carried out for inversions based on the first- and third-derivative regularizations, given by Eqs. (10) and (12), respectively. Minimizing the first derivative produces the flattest sonobuoy tracks, which deviate from the initial positions as little as possible. The results for this case, given in Fig. 7(a), show reasonably small uncertainty ellipses for the first few sonobuoy positions along the track. However, for subsequent sonobuoy positions, which are farther from the initial

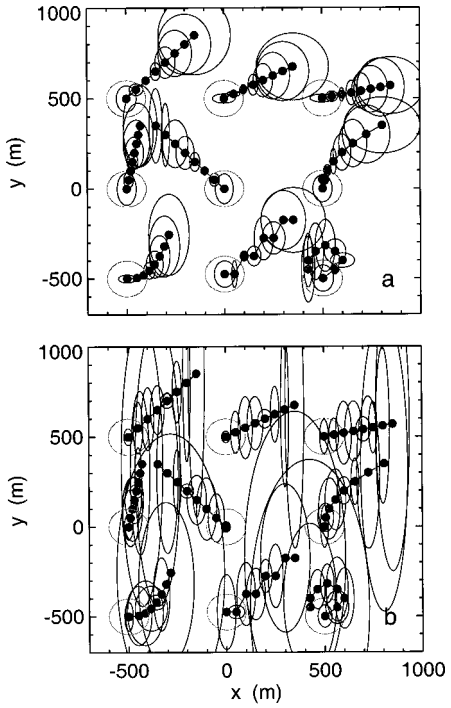


FIG. 7. Results of Monte Carlo appraisal analysis for sonobuoy tracks with minimum first and third derivatives are shown in (a) and (b), respectively. Data errors are 1 ms, prior estimate errors are 50 m for source positions, 100 m for initial sonobuoy positions (gray circles), and 5 m for source and sensor depths.

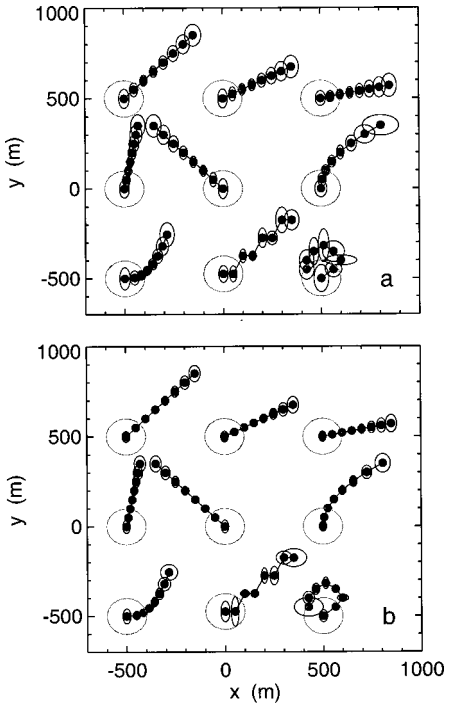


FIG. 8. Results of Monte Carlo appraisal analysis for the smoothest sonobuoy tracks using direct-arrival data. Data errors are 10 ms in (a) and 0.1 ms in (b); prior estimate errors are 50 m for source positions, 100 m for initial sonobuoy positions (gray circles), and 5 m for source and sensor depths.

position, the uncertainties increase substantially. Overall, the results are considerably poorer than those for the smoothest tracks shown in Fig. 5, and the average and maximum positioning errors are substantially larger (Table I).

Minimizing the third derivative of the sonobuoy tracks favors tracks that follow quadratic paths. The appraisal results obtained for this case, shown in Fig. 7(b), are very poor, particularly for the later sonobuoy positions along the tracks. It is interesting to note that even though the sonobuoy positions from the individual inversions of the Monte Carlo appraisal differ substantially from the true positions (e.g., a maximum positioning error of 849 m in Table I), the tracks nonetheless fit the acoustic data to within the χ^2 criterion. As described in Sec. II A, minimizing the third derivative leaves three constants of integration to be determined by the data, and provides weaker *a priori* information than minimizing lower-order derivatives. This lack of information is likely the cause of the poor results for the third-derivative regularization. In particular, the individual inversions of the Monte Carlo appraisal appeared to be unstable, often requiring ten or more iterations for the model change Δ to decrease below the convergence threshold after the data were fit according to the χ^2 criterion. In a number of cases, the inversion did not converge at all [these cases were omitted from the results in Fig. 7(b)], a situation that was never observed when the first- or second-derivative regularization was applied. Hence, it appears that for the sonobuoy tracking problem, the third-derivative regularization simply does not provide sufficient *a priori* information for a meaningful solution.

C. Error study

This section considers a number of factors that affect the accuracy of the sonobuoy-tracking inversion, including the uncertainties of the data and of the *a priori* parameter estimates, and the geometric configuration of the acoustic sources. The effect of errors on the travel-time data is considered in Fig. 8. Figure 8(a) shows the results of the Monte Carlo appraisal analysis when random Gaussian errors with a standard deviation of 10 ms are added to the data (i.e., data errors a factor of 10 times larger than those of the canonical test case of Fig. 5). In comparing the results of Fig. 8(a) and Fig. 5, it is apparent that the larger data errors lead to larger uncertainties for the sonobuoy positions; however, the difference is not dramatic. Figure 8(b) shows the appraisal results obtained when Gaussian errors with standard deviation of 0.1 ms are applied to the data (i.e., errors a factor of 10 times smaller than those of the canonical test case). The sonobuoy positioning errors in Fig. 8(b) and Fig. 5 are very similar, indicating that smaller data errors do not lead to significantly smaller sonobuoy positioning errors. The results in Fig. 8 and Table I indicate that travel-time data errors are not the limiting factor in the sonobuoy tracking inversion.

The effect of errors in the *a priori* parameter estimates is considered in Fig. 9. Figure 9(a) shows the Monte Carlo appraisal results when the uncertainties in the prior estimates of the source and initial sonobuoy positions are twice those of the canonical test case (i.e., 100 m in source positions, 200 m in initial sonobuoy positions, and 10 m in depth for all sources and sensors). Comparing the results of Fig. 9(a) and Fig. 5 shows that the larger uncertainties in the prior parameter estimates lead to substantially larger uncertainties for the

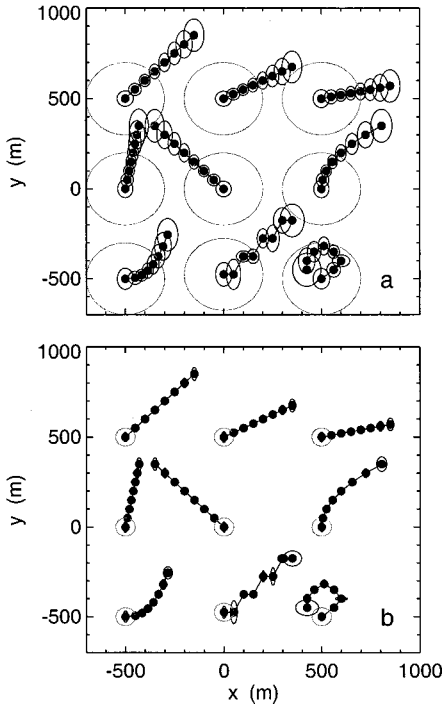


FIG. 9. Results of Monte Carlo appraisal analysis for the smoothest sonobuoy tracks using direct-arrival data with 1 ms errors. In (a) prior estimate errors are 100 m for source positions, 200 m for initial sonobuoy positions (gray circles), and 10 m for source and sensor depths; in (b) prior estimate errors are 25 m for source positions, 50 m for initial sonobuoy positions (gray circles), and 2.5 m for source and sensor depths.

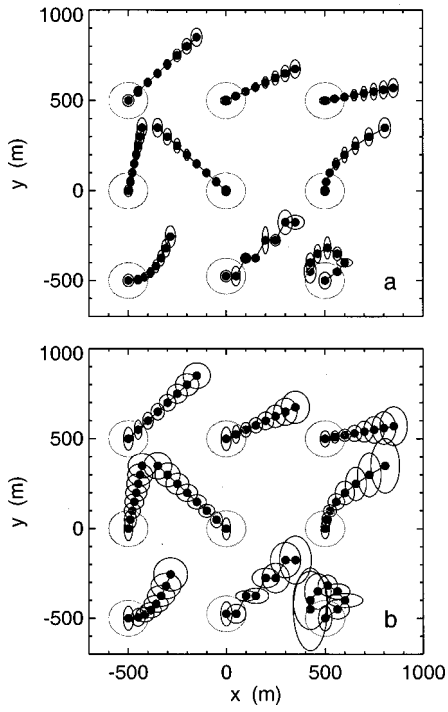


FIG. 10. Results of Monte Carlo appraisal analysis for the smoothest sonobuoy tracks using direct-arrival data with 1 ms errors. In (a) the eight sources are distributed in a circular configuration of radius 2 km. In (b) the sources are distributed in a semicircular configuration in the top half of the coordinates system.

sonobuoy positions. Figure 9(b) shows the appraisal results obtained when the uncertainties for the prior estimates are half those of the canonical case (i.e., 25 m in source positions, 50 m in initial sonobuoy positions, and 2.5 m in depth for all sources and sensors). In this case, the sonobuoy position uncertainties are considerably smaller than those in Fig. 5. The results of Fig. 9 and Table I indicate that the accuracy of the prior parameter estimates is an important factor controlling the uncertainty of the inversion results in the sonobuoy tracking problem. The results of Figs. 8 and 9 indicate that, in practice, there is more to be gained by improving the prior estimates (e.g., improved deployment positioning) rather than measuring arrival times to exceedingly high precision.

The geometric configuration of the acoustic source positions is another factor affecting the uncertainty of the sonobuoy positions. The effect of the source configuration is investigated in Fig. 10. Figure 10(a) shows the sonobuoy track uncertainties computed for a set of eight acoustic sources spaced at 45° angular increments in a circular configuration with a radius of 2 km about the coordinate system origin. The sizes of the error ellipses obtained for the circular configuration are very similar to those computed for the canonical test case (Fig. 5), which involved eight sources located on the $\pm x$- and $\pm y$-axes (Fig. 1). These two source configurations also produced almost identical average and maximum error values (Table I). In fact, similar results were obtained for a number of other source configurations (not shown) which surrounded the sonobuoy field. Symmetry in the source configuration is not required; rather, what is important is that source–receiver ray paths exist with significant components along the $\pm x$- and $\pm y$-directions to provide adequate control in all directions. Figure 10(b) shows the sonobuoy track uncertainties computed for eight sources spaced at 22.5° angular increments in a semicircular configuration with a radius of 2 km in the upper hemisphere of the coordinate system. In this case, the sources do not surround the sonobuoy field, and the resulting error ellipses are significantly larger than for the circular configuration [Fig. 10(a)] or the canonical test case (Fig. 5, Table I). Other source configurations considered (not shown) which did not surround the sonobuoy field produced similar and poorer results. For instance, deploying all sources at one location or along a straight line leads to an unstable inversion that precludes meaningful tracking results. Hence, employing a suitable configuration of sources is an important factor in accurate sonobuoy-track inversion.

The final factor affecting the uncertainty of the sonobuoy tracks is the number of source transmissions, and hence the number of sonobuoy positions that are estimated along each track. A minimum of four transmissions from distinct source locations around the sonobuoy field is required for localization using relative travel-time data. Localization with the minimum number of sources typically gives the poorest results in array element localization, with significant improvements obtained as more sources are employed.[14] Figure 11 shows the results of using the first four source transmissions to localize the first four sonobuoy positions of the canonical test case. The results are indeed significantly poorer that those obtained with eight sources (Fig. 5); however, the tracks obtained with an average position error of 78

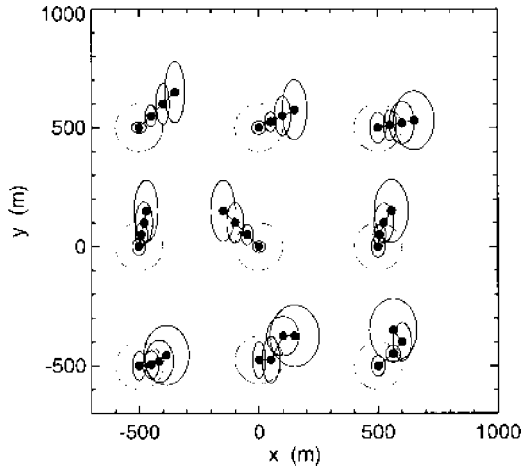


FIG. 11. Results of Monte Carlo appraisal analysis for the smoothest sonobuoy tracks using direct-arrival data from only four sources located on the coordinate axes at ± 3000 m.

m (Table I) might still represent a useful intermediate result until more sources are deployed to refine the estimates.

IV. SUMMARY AND DISCUSSION

This paper considered the problem of localizing a field of freely drifting sonobuoys using relative travel-time measurements from nonsimultaneous acoustic sources deployed about the field for multistatic sonar applications. This represents a particularly challenging inverse problem, since significant independent sonobuoy motion can occur between successive source transmissions, and the source positions themselves are only approximately known. The inversion algorithm developed here solves for the track of each sonobuoy (parametrized by the sonobuoy position at the time of each source transmission), allowing arbitrary sonobuoy motion between source transmissions, as well as for source positions and transmission instants. This leads to a strongly underdetermined inverse problem, with more than three times as many unknown parameters as data. For such problems, an infinite number of solutions exist which fit the data. The method of regularization was applied to formulate a unique inversion for a physically meaningful solution by incorporating *a priori* information regarding the solution and/or its derivatives. The *a priori* information consisted of prior estimates (with uncertainties) for the source positions and initial sonobuoy positions, and a physical model for preferred sonobuoy motion. Three models for sonobuoy motion were considered, consisting of minimizing the first, second, or third spatial derivative of the sonobuoy track. These approaches produce the sonobuoy tracks that deviate minimally from the initial positions, from a smooth function, or from a quadratic function, respectively, subject to fitting the acoustic data to a statistically appropriate level. The smoothest (minimum-curvature) solution is physically reasonable, and was found to perform significantly better than the others on synthetic test cases.

A number of aspects of the sonobuoy-tracking problem were considered by applying the inversion algorithm to variations on a canonical test case. The canonical case involved eight sources deployed around a field of nine sonobuoys which followed independent linear, curved, and zigzag tracks. The data consisted of direct-path arrival times with Gaussian errors of standard deviation 1 ms, and the *a priori* information included prior estimates of the source positions (with random errors of 50-m standard deviation), initial sonobuoy positions (100-m errors), and source and sonobuoy depths (5-m errors). Comparisons between the canonical test case and other cases were made in terms of a Monte Carlo appraisal analysis which provided uncertainty ellipses for sonobuoy positions along the tracks. This analysis indicated that including travel times of bottom-reflected arrivals in the data set provides little benefit for sonobuoy tracking if the receiver depths are relatively well known, since these arrivals follow the same horizontal paths as the direct arrivals. It was found that varying the uncertainties of the travel-time data from 0.1–10 ms did not strongly affect the uncertainties of the sonobuoy positions. However, increasing or decreasing the uncertainties of the prior parameter estimates by a factor of 2 had a substantial effect on tracking accuracy, indicating that it is the accuracy of the prior estimates, and not of the data, that limits the accuracy of sonobuoy tracking in this test case. The geometric configuration of the source positions also affects sonobuoy tracking. The best results were obtained with source configurations that surrounded the sonobuoy field, providing source–receiver ray paths with significant components in the $\pm x$- and $\pm y$-directions. Finally, the number of source transmissions (and hence, the number of sonobuoy positions along the track) affects the average positioning accuracy. The tracking results obtained using the minimum number of source transmissions (four) were significantly poorer than those for the canonical test case, but represent reasonable first estimates that could be improved upon as further transmissions were recorded.

The approach developed in this paper for tracking sonobuoys is general, and could be applied to other localization problems involving moving acoustic receivers (or sources). The examples presented here represent a particularly challenging scenario; significantly smaller localization errors can be expected in cases where the time interval between successive source transmissions is small and/or if source transmission instants are independently known, allowing the measured data to be interpreted as absolute travel times.[11]

APPENDIX: RAY THEORY IMPLEMENTATION

For completeness, this Appendix describes the implementation of ray theory applied to compute the acoustic travel times and partial derivatives required by the inversion algorithm described in Sec. II. Consider an acoustic source and receiver in the ocean at (x_i, y_i, z_i) and (x_j, y_j, z_j), respectively, with $z_j < z_i$ (for the reverse, z_i and z_j should be interchanged in the following equations). The horizontal range between source and receiver is given by

$$r = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \tag{A1}$$

Expressions for the range r and arrival time t along a (nonturning) ray path between source and receiver in an

ocean with sound-speed profile $c(z)$ are easily derived by applying Snell's Law to an infinite stack of infinitesimal layers[15]

$$r=\int_{z_j}^{z_i}\frac{pc(z)dz}{[1-p^2c^2(z)]^{1/2}}, \tag{A2}$$

$$t=t_0+\int_{z_j}^{z_i}\frac{dz}{c(z)[1-p^2c^2(z)]^{1/2}}, \tag{A3}$$

where t_0 represents the source instant. In (A1) and (A2), the ray parameter $p=\cos\theta(z)/c(z)$ is constant along a ray path, and defines the take-off (grazing) angle at the source. The ray parameter for an eigenray connecting source and receiver is usually determined by searching for the value of p which produces the correct range (to a specified tolerance) using (A2). An efficient procedure of determining p for direct-path eigenrays is based on Newton's method.[10] An initial estimate p_0 is calculated assuming straight-line propagation with a sound speed $\overline{c}$ representing the harmonic mean of the measured sound-speed profile between source and receiver

$$\overline{c}=(z_i-z_j)\Big/\int_{z_j}^{z_i}\frac{dz}{c(z)}. \tag{A4}$$

An improved estimate p_1 is obtained by expanding $r(p)$ in a Taylor's series about p_0 and neglecting nonlinear terms to give

$$r(p)=r(p_0)+\frac{\partial r(p_0)}{\partial p}(p_1-p_0), \tag{A5}$$

which has a solution

$$p_1=p_0+\left[\frac{\partial r(p_0)}{\partial p}\right]^{-1}(r(p)-r(p_0)). \tag{A6}$$

In (A6), $\partial r/\partial p$ is determined by differentiating (A2) to yield

$$\frac{\partial r}{\partial p}=\int_{z_j}^{z_i}\frac{c(z)dz}{[1-p^2c^2(z)]^{3/2}}. \tag{A7}$$

If $r(p_1)$ computed from (A2) is within the tolerance of the desired range, the procedure is complete. If not, the starting value is updated, $p_0\leftarrow p_1$, and the procedure repeated iteratively until a satisfactory value is obtained. The travel time along the ray path is then computed using (A3). Since Newton's method converges quadratically near the solution, this is an efficient method of determining direct eigenrays to high precision.

In addition to computing travel times, the linearized inversion algorithm requires partial derivatives of travel time with respect to source and receiver coordinates. Consider first the partial derivative with respect to horizontal coordinate x_i. Employing the chain rule

$$\frac{\partial t}{\partial x_i}=\frac{\partial t}{\partial p}\frac{\partial p}{\partial r}\frac{\partial r}{\partial x_i}=\frac{\partial t}{\partial p}\left[\frac{\partial r}{\partial p}\right]^{-1}\frac{\partial r}{\partial x_i}. \tag{A8}$$

The three partials on the right side of (A8) can be calculated from (A3), (A2), and (A1), respectively, yielding

$$\frac{\partial t}{\partial x_i}=p(x_i-x_j)/r. \tag{A9}$$

Similarly, partial derivatives with respect to the other horizontal coordinates are

$$\frac{\partial t}{\partial x_j}=p(x_j-x_i)/r, \tag{A10}$$

$$\frac{\partial t}{\partial y_i}=p(y_i-y_j)/r, \tag{A11}$$

$$\frac{\partial t}{\partial y_j}=p(y_j-y_i)/r. \tag{A12}$$

The partial derivative of t with respect to vertical coordinate z_i can be determined by differentiating (A3) to give

$$\frac{\partial t}{\partial z_i}=\int_{z_j}^{z_i}\frac{pc(z)dz}{[1-p^2c^2(z)]^{3/2}}\left(\frac{\partial p}{\partial z_i}\right)-\frac{1}{c(z_i)[1-p^2c^2(z_i)]^{1/2}}. \tag{A13}$$

An expression for $\partial p/\partial z_i$ can be obtained by noting that

$$\frac{\partial r}{\partial z_i}=0=\int_{z_j}^{z_i}\frac{c(z)dz}{[1-p^2c^2(z)]^{3/2}}\left(\frac{\partial p}{\partial z_i}\right)-\frac{pc(z_i)}{[1-p^2c^2(z_i)]^{1/2}}. \tag{A14}$$

Solving for $\partial p/\partial z_i$ and substituting into (A13) yields

$$\frac{\partial t}{\partial z_i}=\frac{1}{c(z_i)}[1-p^2c^2(z_i)]^{1/2}. \tag{A15}$$

Similarly

$$\frac{\partial t}{\partial z_j}=-\frac{1}{c(z_j)}[1-p^2c^2(z_j)]^{1/2}. \tag{A16}$$

Finally, the partial derivative with respect to the source instant from (A3) is

$$\frac{\partial t}{\partial t_0}=1. \tag{A17}$$

To implement the above equations, it is assumed that a discrete sound-speed profile can be represented by a series of layers with a linear sound-speed gradient in each layer. The simplest ray paths to trace involve reflections from a flat sea bottom, since these do not involve turning points (i.e., points where the ray passes through zero grazing angle and changes vertical direction as the result of refraction). Sea-bottom reflections are modeled using the method of images, i.e., representing the reflected path by a direct ray path from an image source located below the bottom.[15] To apply the method of images, the sound-speed profile is reflected about the interface in the same manner as the sources. In the following, let $\{(z_k,c_k)\}$ represent the sound-speed profile including the requisite reflection, $\{c_k'\}$ be the corresponding sound speed gradients, and z_i and z_j be the source and receiver depths, respectively. For the case of linear sound-speed gradients, the integrals in Eqs. (A2), (A3), and (A7) can be evaluated analytically, yielding the following results, where $w_k\equiv(1-p^2c_k^2)^{1/2}$:

$$r=\sum_{k=j}^{i-1}\frac{w_k-w_{k+1}}{pc_k'}, \tag{A18}$$

$$t=t_0+\sum_{k=j}^{i-1}\frac{1}{c_k'}\left[\log_e\frac{c_{k+1}(1+w_k)}{c_k(1+w_{k+1})}\right], \tag{A19}$$

$$\frac{\partial r}{\partial p}=\sum_{k=j}^{i-1}\frac{w_k-w_{k+1}}{p^2c_k'w_kw_{k+1}}. \tag{A20}$$

Calculating integrals along direct ray paths is somewhat more complicated, since these paths can involve turning rays. Rather than simply integrating (summing) along the ray path as per the reflected rays above, the possibility that the direct ray turns within a specific layer must be checked as it enters each layer along its path. Consider the case of an upward-propagating ray entering the lth layer. The turning depth for this ray is given by

$$z_T = z_{l+1} - (1/p - c_{l+1})/c_l' . \tag{A21}$$

If this depth is less than z_l (the top of the lth layer) the ray turns: if not, it proceeds into layer $l-1$. If the direct ray does not turn between source and receiver, (A18)–(A20) apply. However, if the direct ray turns in layer l, then four steps are required: (i) integrate from z_j to z_{l+1} (the bottom of the lth layer) using the above equations; (ii) integrate from z_{l+1} to the turning depth z_T (where, by definition, $c_T = 1/p$ and $w_T = 0$); (iii) integrate from z_T to z_{l+1}; and (iv) integrate from z_{l+1} to z_i. Applying this procedure leads (after some algebra) to the following equations for turning rays:

$$r = \sum_{k=j}^{l+2} \frac{w_k - w_{k-1}}{p c_{k-1}'} - \frac{2 w_{l+1}}{p c_l'} + \sum_{k=l+1}^{i-1} \frac{w_k - w_{k+1}}{p c_k'}, \tag{A22}$$

$$t = t_0 + \sum_{k=j}^{l+2} \frac{1}{c_{k-1}'} \left[\log_e \frac{c_{k-1}(1+w_k)}{c_k(1+w_{k-1})} \right] + \frac{2}{c_l'} \log_e \frac{p c_{l+1}}{1+w_{l+1}} + \sum_{k=l+1}^{i-1} \frac{1}{c_k'} \left[\log_e \frac{c_{k+1}(1+w_k)}{c_k(1+w_{k+1})} \right], \tag{A23}$$

$$\frac{\partial r}{\partial p} = \sum_{k=j}^{l+2} \frac{w_k - w_{k-1}}{p^2 c_{k-1}' w_k w_{k-1}} - \frac{2}{c_l' p^2 w_{l+1}} + \sum_{k=l+1}^{i-1} \frac{w_k - w_{k+1}}{p^2 c_k' w_k w_{k+1}} . \tag{A24}$$

[1] P. Nicolas, "Localization of far-field sources with an array of unknown geometry," SACLANT-CEN Report SR-172 (1990).

[2] P. Nicolas and S. Bongi, "Localization of far-field sources with an array of unknown geometry: Experimental results with a field of sonobuoys," SACLANTCEN Report SR-281 (1994).

[3] C. M. McIntyre, J. Wang, and L. Kelly, "The effect of position uncertainty in multistatic acoustic localization," Proceedings of the International Conference on Information, Decision, and Control, Adelaide, Australia (1999).

[4] M. V. Greening and S. J. Lourey, "Sonobuoy field estimation," UDT Pacific 2000 Conference Proceedings, Sydney, Australia, pp. 94–97 (2000).

[5] W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory* (Academic, Orlando, 1984).

[6] S. C. Constable, R. L. Parker, and C. G. Constable, "Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data," Geophysics **52**, 289–300 (1987).

[7] C. van Schooneveld, "Inverse problems: A tutorial survey," in *Underwater Acoustic Data Processing*, edited by Y. T. Chan (Kluwer, The Netherlands, 1989), pp. 393–411.

[8] J. A. Scales, P. Docherty, and A. Gersztenkorn, "Regularization of nonlinear inverse problems: Imaging the near-surface weathering layer," Inverse Probl. **6**, 115–131 (1990).

[9] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, New York, 1992).

[10] S. E. Dosso, M. R. Fallat, B. J. Sotirin, and J. L. Newton, "Array element localization for horizontal arrays via Occam's inversion," J. Acoust. Soc. Am. **104**, 846–859 (1998).

[11] S. E. Dosso, G. H. Brooke, S. J. Kilistoff, B. J. Sotirin, V. K. McDonald, M. R. Fallat, and N. E. Collison, "High-precision array element localization of vertical line arrays in the Arctic Ocean," IEEE J. Ocean. Eng. **23**, 365–379 (1998).

[12] S. E. Dosso and N. E. Collison, "Regularized inversion for towed-array shape estimation," in *Inverse Problems in Underwater Acoustics*, edited by M. Taroudakis and G. Makrakis (Springer, Berlin, 2001).

[13] S. E. Dosso and M. Riedel, "Array element localization for towed marine seismic arrays," J. Acoust. Soc. Am. **110**, 955–966 (2001).

[14] S. E. Dosso and B. J. Sotirin, "Optimal array element localization," J. Acoust. Soc. Am. **106**, 3445–3459 (2000).

[15] W. M. Telford, L. P. Geldart, R. E. Sheriff, and D. A. Keys, *Applied Geophysics* (Cambridge University Press, New York, 1976), Chap. 4.

A revised model of the inner-hair cell and auditory-nerve complex

Christian J. Sumner[a)]
Centre for the Neural Basis of Hearing at Essex, Department of Psychology, University of Essex, Colchester CO4 3SQ, United Kingdom

Enrique A. Lopez-Poveda
Centro Regional de Investigación Biomédica, Facultad de Medicina, Universidad de Castilla-La Mancha, Campus Universitario, 02071 Albacete, Spain

Lowel P. O'Mard and Ray Meddis
Centre for the Neural Basis of Hearing at Essex, Department of Psychology, University of Essex, Colchester CO4 3SQ, United Kingdom

(Received 23 April 2001; revised 1 November 2001; accepted 31 December 2001)

A revised computational model of the inner-hair cell (IHC) and auditory-nerve (AN) complex is presented and evaluated. Building on previous models, the algorithm is intended as a component for use in more comprehensive models of the auditory periphery. It combines smaller components that aim to be faithful to physiology in so far as is practicable and known. Transduction between cochlear mechanical motion and IHC receptor potential (RP) is simulated using a modification of an existing biophysical IHC model. Changes in RP control the opening of calcium ion channels near the synapse, and local calcium levels determine the probability of the release of neurotransmitter. AN adaptation results from transmitter depletion. The exact timing of AN action potentials is determined by the quantal and stochastic release of neurotransmitter into the cleft. The model reproduces a wide range of animal RP and AN observations. When the input to the model is taken from a suitably nonlinear simulation of the motion of the cochlear partition, the new algorithm is able to simulate the rate-intensity functions of low-, medium-, and high-spontaneous rate AN fibers in response to stimulation both at best frequency and at other frequencies. The variation in fiber type arises in large part from the manipulation of a single parameter in the model: maximum calcium conductance. The model also reproduces quantitatively phase-locking characteristics, relative refractory effects, mean-to-variance ratio, and first- and second-order discharge history effects. [DOI: 10.1121/1.1453451]

PACS numbers: 43.64.Bt, 43.66.Ba [LHC]

I. INTRODUCTION

The inner-hair-cell (IHC) and auditory-nerve (AN) complex is the primary apparatus for the transduction of basilar membrane (BM) motion into neural signals. Because transduction is an obligatory stage in the auditory pathway, computational models of IHCs are a vital component in any auditory modeling project. Here, we present a revised version of an earlier IHC model (Meddis, 1986, 1988; Meddis *et al.*, 1990) that improves on previous models in terms of the range of phenomena simulated and is more consistent with recent developments in hair-cell physiology.

IHCs have been the subject of modeling studies for at least three decades (e.g., Davis, 1965; Weiss, 1966). The goals of the research have been varied. Many models focus on possible mechanisms for adaptation (e.g., Schroeder and Hall, 1974; Smith and Brachman, 1982; Eggermont, 1985; Meddis, 1986). Others are concerned with the biophysics of hair cells (e.g., Shamma *et al.*, 1986; Rattay *et al.*, 1998) or the mechano-electric transduction process (Corey and Hudspeth, 1983). The purpose of the model to be described below is primarily to generate an accurate representation of the input–output characteristics of the hair cell for arbitrary stimuli. Physiological accuracy, for its own sake, is a lesser concern. However, inclusion of key physiological processes (as far as these are known) is likely to be a productive strategy and this is the approach that is adopted below.

Many older models (e.g., Meddis, 1986; Carney, 1993) simulate the response of high-spontaneous rate (HSR) fibers to pure tones at the fiber's best frequency (BF). However, they make no specific claim to represent the response of medium (MSR) and low (LSR) spontaneous rate fibers, nor do they claim to deal with pure-tone responses for frequencies away from BF. It is clear from AN measurements that the characteristics of MSR and LSR fiber types are strongly influenced by nonlinear aspects of mechanical filtering (Yates *et al.*, 1990). The same nonlinearities can also explain why the shape of rate-intensity (RI) functions for a single AN fiber does vary considerably with stimulating frequency. The new model simulates these effects by including a nonlinear model of the response of the BM as the input to the hair cell. This study joins several recent reports that address the modeling of nonlinear filtering (Robert and Erikson, 1999; Irino and Patterson, 2001; Zhang *et al.*, 2001). How-

[a)]Current address: Kresge Hearing Research Institute, University of Michigan, Ann Arbor, 48109-0506. Electronic mail: cjsumner@umich.edu

ever, we emphasize the development of the IHC model and its rate responses to pure-tone stimuli. The response to complex stimuli will be reported elsewhere. Also, the testing of the model is restricted wherever possible to guinea-pig data at high BFs.

An earlier version of the new model (Lopez-Poveda *et al.*, 1998) used an existing biophysical model of the IHC receptor potential (RP) (Davis, 1965; Shamma *et al.*, 1986), to control the synaptic adaptation stage from the original Meddis IHC model. They simulated the control relationship between the RP and the transmitter release probability as an exponential function, and were able to reproduce the characteristic input/output functions associated with HSR, MSR, and LSR fibers. However, the exponential function led to unrealistically high release rates at high stimulus levels. Schoonhoven *et al.* (1997) employed a similar model but found a saturating exponential to be a more suitable function. Recently, Robert and Erikson (1999) have employed a quadratic function for the same purpose. In the new model, we hypothesize that a single presynaptic calcium current controls the rate of transmitter release and have explicitly modeled this process. Calcium currents are necessary for neurotransmitter release at all chemical synapses (Augustine *et al.*, 1985a). These have the advantage of offering a naturally saturating, exponential function. We find that variations in maximum calcium conductance can be successfully used to change the fiber response from low- through to high-spontaneous rate type of RI function.

The new model uses a previous simulation of the circulation of transmitter at the synapse (Meddis, 1986) but the release of neurotransmitter is now quantal and stochastic. Further, the release of a single quantum of transmitter is assumed to initiate an action potential (AP) in the postsynaptic AN fiber, if it is not already in a refractory state. This idea was suggested by Geisler (1981), and is supported by observations of postsynaptic potentials (Siegel, 1992). Thus, the stochastic nature of AP generation in AN fibers is treated as *synaptic* in origin. This stochastic release leads to first- and second-order discharge history effects. Postsynaptic refractory effects are modeled with a simple absolute refractory period and a short exponential relative refractory period. The model does not include any of the postsynaptic effects associated with the lateral olivo-cochlear efferent system.

The restriction of phase locking in mammals to low-frequency signals is an important characteristic of the AN response (Kiang *et al.*, 1965). Much of the restriction is thought to originate with the low-pass filtering of the IHC membrane (Palmer and Russell, 1986). Unfortunately, the decline of phase locking in the Meddis (1986) IHC was erroneously modeled as an exclusive function of transmitter recycling processes. Moreover, some phase locking in the model persisted above 5 kHz. Some models have addressed this issue with simple low-pass filtering (Carney, 1993; Robert and Erikson, 1999), while others (Schoonhoven *et al.*, 1997; Lopez-Poveda *et al.*, 1998) have adapted biophysical models of membrane filtering. Presynaptic calcium channels have also been implicated in the restriction of AN phase locking (Kidd and Weiss, 1990). In the present model, phase locking is restricted to low frequencies by a combination of low-pass filtering of the RP, the dynamics of calcium flow, and transmitter recycling.

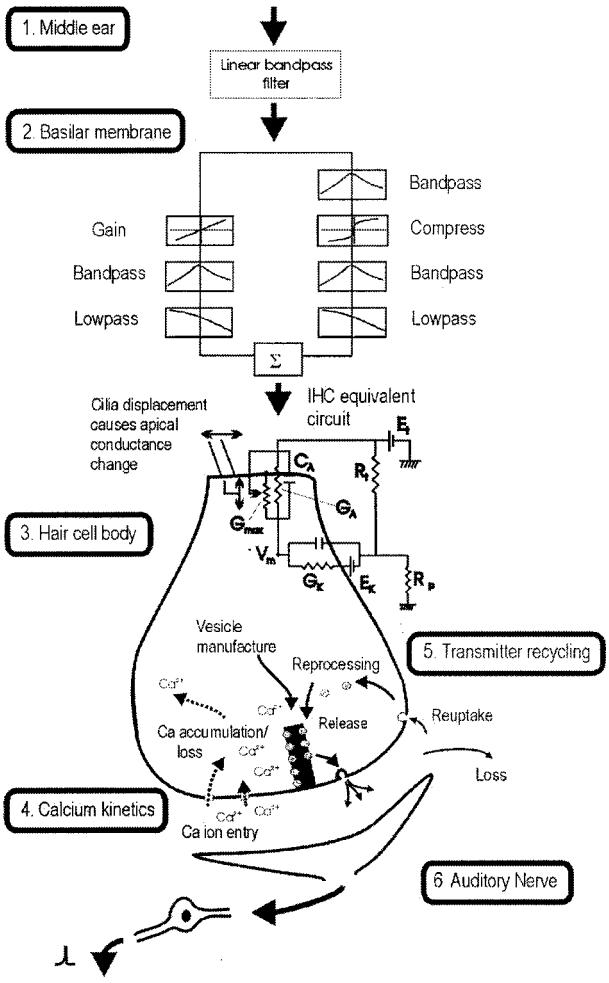


FIG. 1. Schematic diagram of the complete composite auditory model.

II. THE IHC MODEL

The input to the IHC model is motion of the cochlear partition, and its output is a stream of transmitter release events. However, a full evaluation of the model requires assumptions concerning the middle and inner ear as well as the functioning of AN fibers. The details of the IHC model are described in this section, while the supportive stages required for modeling are presented in Sec. III. Figure 1 shows the complete processing path for the model, from sound entering the auditory canal to spiking on the auditory nerve.

A. IHC receptor potential

The model incorporates a modified version of the simulation by Shamma *et al.* (1986) of the transduction of BM motion into RP. The displacement of the IHC cilia, $u(t)$, as a function of BM velocity, $v(t)$, is given by

$$\tau_c \frac{du(t)}{dt} + u(t) = \tau_c C_{\text{cilia}} v(t), \tag{1}$$

where C_{cilia} is a gain factor and τ_c is a time constant. Thus, cilia move in phase with BM velocity at low frequencies and with displacement at high frequencies. The cilia displace-

ment causes a change in the number of open ion channels and, consequently, in the apical conductance $G(u)$. The total apical conductance is given by

$$G(u)=G_{\text{cilia}}^{\max}\left[1+\exp\left(-\frac{u(t)-u_0}{s_0}\right)\times\left[1+\exp\left(-\frac{u(t)-u_1}{s_1}\right)\right]\right]^{-1}+G_a, \quad (2)$$

where $G_{\text{cilia}}^{\max}$ is the transduction conductance with all channels open, and G_a is the passive conductance in the apical membrane. The proportion of open channels is modeled as a three-state energy barrier (Boltzmann) function (see Mountain and Hubbard, 1996), where s_0, u_0, s_1, and u_1 are constants determining the exact shape of the nonlinearity. These are chosen to fit cilia displacement-depolarization functions (Russell *et al.*, 1986). Equation (2) replaces a first-order Boltzmann function employed by Shamma *et al.* The membrane potential of the cell body is modeled with a passive electrical circuit analog (Fig. 1, module 3), described by

$$C_m\frac{dV(t)}{dt}+G(u)(V(t)-E_t)+G_k(V(t)-E_k')=0, \quad (3)$$

where $V(t)$ is the intracellular hair cell potential; C_m is the cell capacitance; G_k is the voltage-invariant basolateral membrane conductance; E_t is the endocochlear potential; and $E_k'=E_k+E_tR_p/(R_t+R_p)$ is the reversal potential of the basal current E_k (mostly potassium; Corey and Hudspeth, 1983), corrected for the resistance (R_t,R_p) of the supporting cells (see Shamma *et al.*, 1986).

B. Calcium controlled transmitter release function

The release of neurotransmitter into the synaptic cleft is mediated by calcium ions in all synapses. Augustine *et al.* (1985a) showed that postsynaptic potentials were only observed when presynaptic depolarization coincided with the presence of extracellular Ca^{2+} (see Johnston and Wu, 1995, for a review). The transmitter release function is a three-part process, labeled as module 4 in Fig. 1.

1. Depolarization of the IHC membrane leads to the opening of calcium ion channels

We employ a third-order process similar to Hudspeth and Lewis (1988) and Kidd and Weiss (1990), but with a time constant that is not voltage dependent. Calcium current (I_{Ca}) is a function of the RP

$$I_{\text{Ca}}(t)=G_{\text{Ca}}^{\max}m_{I_{\text{Ca}}}^3(t)(V(t)-E_{\text{Ca}}), \quad (4)$$

where E_{Ca} is the reversal potential for calcium and $G_{\text{Ca}}^{\max}$ is the calcium conductance in the vicinity of the synapse, with all the channels open. $m_{I_{\text{Ca}}}(t)$ is the fraction of calcium channels that are open. The steady-state value of the latter, $m_{I_{\text{Ca}},\infty}$, is modeled by a Boltzmann function

$$m_{I_{\text{Ca}},\infty}=[1+\beta_{\text{Ca}}^{-1}\exp(\gamma_{\text{Ca}}V(t))]^{-1}, \quad (5)$$

where β_{Ca} and γ_{Ca} are constants chosen to reflect published observations of calcium currents (see Table II), and $m_{I_{\text{Ca}}}(t)$ is a low-pass-filtered function of $m_{I_{\text{Ca}},\infty}$

$$\tau_{I_{\text{Ca}}}\frac{dm_{I_{\text{Ca}}}(t)}{dt}+m_{I_{\text{Ca}}}(t)=m_{I_{\text{Ca}},\infty}, \quad (6)$$

where $\tau_{I_{\text{Ca}}}$ is a time constant.

2. Calcium ions enter the cell, and accumulate briefly in the vicinity of the synapse

Calcium concentration $[Ca^{2+}](t)$ is modeled as a first-order, low-pass-filtered function of calcium current, $I_{\text{Ca}}(t)$ (after Hudspeth and Lewis, 1988)

$$\tau_{[\text{Ca}]}\frac{d[Ca^{2+}](t)}{dt}+[Ca^{2+}](t)=I_{\text{Ca}}(t), \quad (7)$$

where $\tau_{[\text{Ca}]}$ is a time constant.

3. The probability of the release of transmitter is proportional to the cube of Ca^{2+} concentration

We employ the function

$$k(t)=\max(([Ca^{2+}]^3(t)-[Ca^{2+}]_{\text{thr}}^3)z,0), \quad (8)$$

where $[Ca^{2+}]_{\text{thr}}$ is a threshold constant, z is a scalar for converting calcium concentration levels into release rate, and the cube function is based on data by Augustine *et al.* (1985b).

C. Quantal and probabilistic model of synaptic adaptation

The transmitter release rate, $k(t)$, drives a model of synaptic adaptation identical to that given in Lopez-Poveda *et al.* (1998). It simulates the functional characteristics of adaptation, which are assumed here to be due to presynaptic transmitter depletion. The scheme is the same as that proposed by Meddis (1986) except that release of transmitter into the cleft is now quantal and stochastic. It is described by the following equations:

$$\frac{dq(t)}{dt}=N(w(t),x)+N([M-q(t)]y)-N(q(t),k(t)), \quad (9)$$

$$\frac{dc(t)}{dt}=N(q(t),k(t))-lc(t)-rc(t), \quad (10)$$

$$\frac{dw(t)}{dt}=rc(t)-N(w(t),x). \quad (11)$$

Figure 1, module 5, shows one possible physical interpretation of this system. Individual vesicles of neurotransmitter (probably glutamate), are released from the *immediate* (q) store into the *cleft* (c), at a rate, $k(t)$, that is dependent on calcium concentration. In the cleft, the transmitter disperses and some is lost from the system at a rate l. The remaining transmitter in the cleft is taken back into the cell into a *reprocessing* (w) store at a rate r. Here, it is repackaged into vesicles that are returned to the immediate store at a rate x. Additionally, q is continuously replenished with new transmitter vesicles at a rate, $y[M-q(t)]$, where M represents the maximum number of transmitter quanta that can be held in the immediate store (q).

Neurotransmitter in the immediate store is quantal, and enters and leaves stochastically. The stochastic transport of neurotransmitter is described by the function $N(n,\rho)$, in which each of n quanta has an equal probability of release, ρdt, in a single simulation epoch. In the cleft and reprocessing stores, transmitter is a continuous quantity. This means, for instance, that the contents of the reprocessing store must be an integer number greater than 1 for a transmitter quantum to be eligible to rejoin the immediate store. The output from the synapse is a stream of discrete events indicating vesicle releases, $N(q(t),k(t))$. More detailed accounts of this process in a nonstochastic form can be found in Meddis (1986, 1988), Meddis *et al.* (1991) and Hewitt and Meddis (1991).

III. INPUT TO AND OUTPUT FROM THE IHC

Very few studies apply stimuli directly to, or record directly from, IHCs. As a consequence, full evaluation of our IHC algorithm requires that we simulate both the proximal stimulus (BM velocity) in response to acoustic stimuli and the typical response in AN fibers. To do this we need supplementary models. In our evaluation we have used a model of the low- and high-frequency attenuation of the middle ear (Sec. III A) and a nonlinear filtering operation to represent the motion of the BM (Sec. III B). Section III C describes the procedure used to simulate the conversion of release events into APs in the AN itself.

A. Middle ear

The response of middle ear is modeled by a second-order linear bandpass Butterworth filter with an upper cutoff of 22 kHz and a lower cutoff of 12.5 kHz. This was to fit to data from a single guinea-pig middle ear at high frequencies (Nuttall and Dolan, 1996). The input to the filter is sound pressure (μPa) and the output is stapes velocity, $x(t)$, in m/s. This transformation was achieved with a scale factor of 1.4 $\times 10^{-10}$ ms^{-1} μPa^{-1}.

B. DRNL cochlear filtering

The filtering of the BM is modeled with a "dual-resonance-nonlinear" (DRNL) filter architecture which has been described and evaluated more fully elsewhere (Meddis *et al.*, 2001). The input to the DRNL is stapes velocity and the output is BM velocity, $v(t)$. Figure 1, module 2, shows the basic architecture. A single DRNL filter consists of two parallel pathways, one linear and the other nonlinear, whose outputs are summed to produce the filter output. The linear pathway is a gammatone filter (Patterson *et al.*, 1988) followed by a low-pass filter and a gain. The nonlinear pathway consists of the following cascade: a gammatone filter, a compression function, a second gammatone filter, and a low-pass filter. The compression in the nonlinear pathway is described by

$$y[t]=\text{sign}(x[t])\times\min(a|x[t]|,b|x[t]|^{\nu}), \tag{12}$$

where a, b, and ν are parameters of the model that vary as a function of location along the partition. The DRNL parameters are initially fitted to guinea-pig BM laser-interferometry data at a site with a BF of 18 kHz (Nuttall and Dolan, 1996), as described by Meddis *et al.* (2001).

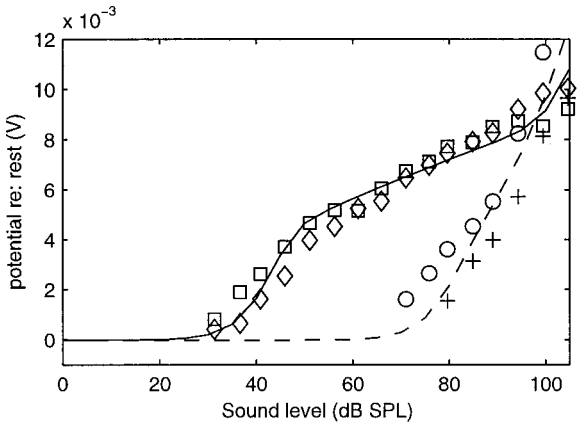


FIG. 2. dc receptor potential. Unconnected symbols are animal data recorded from IHC with 18-kHz BF (Patuzzi and Sellick, 1983), using 7-kHz stimulus (+, ○) and 18-kHz stimulus (□, ◇). Continuous lines are the model response at 7 kHz (dashed) and at 18 kHz (solid).

C. AN response

Postsynaptic intracellular recordings from the dendrites of guinea-pig AN bipolar neurons suggest that the release of a single quantum of transmitter is sufficient to produce an AP in an AN fiber, at least when the fiber is in a nonrefractory state (Siegel, 1992). The input to the AN stage of the model is a stream of transmitter release events from the synapse. Refractoriness (both absolute and relative) is simulated by a uniform random process. For each vesicle release, an AP is generated in the fiber if a vesicle is released and $p(t)$ exceeds a random number (0–1)

$$P(t)=0, \quad \text{for } (t-t_l)<R_A \tag{13}$$
$$p(t)=1-c_r e^{-(t-t_l-R_A)/s_r}, \quad \text{for } (t-t_l)\geq R_A,$$

where c_r is 0.55 and determines the maximum contribution of the relative refractory period in $p(t)$; s_r is 0.8 ms, and is the time constant of refraction; t is time now, t_l is the time of the last spike, and R_A is 0.75 ms, the absolute refractory period.

IV. MODEL EVALUATION

The implementation of the complete peripheral model described in Secs. II and III was fitted to high-BF guinea-pig data, at the level of the RP and the AN. All computations were by numerical integration and used a time step of 10 μs. Since the model is inherently stochastic, repeated runs (between 20 and 500) are required to build up reliable post-stimulus time histograms (PSTHs). The implementation was developed using the Development System for Auditory Modeling (DSAM) C libraries.[1]

A. Average rate responses

1. dc component of the receptor potential

The average response to high-frequency stimulation in the IHC is measured as the dc component of the intracellular potential. Unconnected symbols in Fig. 2 show the dc RP of

TABLE I. IHC receptor potential [Eqs. (1), (2) and (3)].

Et, endocochlear potential (V)	$100E-3$
Ek, potassium reversal potential (V)	$-70.45E-3$
*G*0, resting conductance (S=Siemens)	$1.974E-9$
Gk, potassium conductance (S)	$1.8E-8$
E_k correction, $Rp/(Rt+Rp)$	0.04
$G_{\text{cilia}}^{\max}$, max. mechanical conductance (S)	$8E-9$
s_0, displacement sensitivity (m^{-1})	$85E-9$
u_0, displacement offset (m)	$7E-9$
s_1, displacement sensitivity (m^{-1})	$5E-7$
u_1, displacement offset (m)	$7E-9$
C_m, total capacitance (F)	$6E-12$
τ_c, cilia/BM time constant (s)	$2.13E-3$
C_{cilia}, cilia/BM coupling gain (dB)	16

the guinea-pig IHC (Patuzzi and Sellick, 1983), in response to pure-tone stimulation at BF (18 kHz) and off-BF (7 kHz). The two functions have a different shape. This is likely to reflect differences between the compressed response of the BM to stimulation at BF and linear response at frequencies well below BF. The solid and dashed lines show the response of the model at the two stimulus frequencies. The threshold of the animal BF response is higher than often found in AN measurements, which may indicate a loss of tip sensitivity in the animal preparation. To fit these responses, at both stimulus frequencies, required different DRNL parameters from those used by Meddis *et al.* (2001). The new parameters are given in Table III as ''IHC set.'' With an appropriate choice of IHC parameters, cilia coupling gain, C_{cilia}, and mechanically driven conductance, $G_{\text{cilia}}^{\max}$ [Eqs. (1) and (2)], it was possible to engineer an almost-perfect fit between the animal and model data. These are given in Table I.

2. Rate-intensity functions of the AN

In guinea pigs, RI functions at BF have been divided into three categories (Winter *et al.*, 1990). HSR fibers display activity in the absence of any stimulus (SR >18 spikes/s), have a low acoustic threshold ($<$20-dB SPL in the data used here), and a steep RI function at BF that almost completely saturates within 20–30 dB of the threshold. MSR fibers (SR<18 spikes/s) display a ''sloping saturation'' type rate of growth at BF. LSR fibers have almost no spontaneous activity, have high thresholds, and display an RI function at BF with a shallow slope. These different shapes are thought to reflect the linear and compressive regions of BM response (Yates *et al.*, 1990). Figure 3 compares model results with two AN RI functions at BF for each fiber type (from Winter *et al.*, 1990). They all use the same DRNL parameter set (''AN set'' Table III; see below for their derivation). The RP parameters used were the same as in Sec. IV A 1. Only three transmitter release parameters, $G_{\text{Ca}}^{\max}$, $[\text{Ca}^{2+}]_{\text{thr}}$, and *M*, were varied to produce the six model functions (see Table II, columns H_1, H_2, M_1, M_2, L_1, L_2).

It was a weakness of earlier models (e.g., Meddis and Hewitt, 1991; Carney, 1993) of hair-cell functioning that they were unable to model the change in shape of the RI function when frequencies different from BF were used. Figures 4(A) and (C) show examples of RI functions measured both at BF (16 kHz), and at frequencies above and below BF

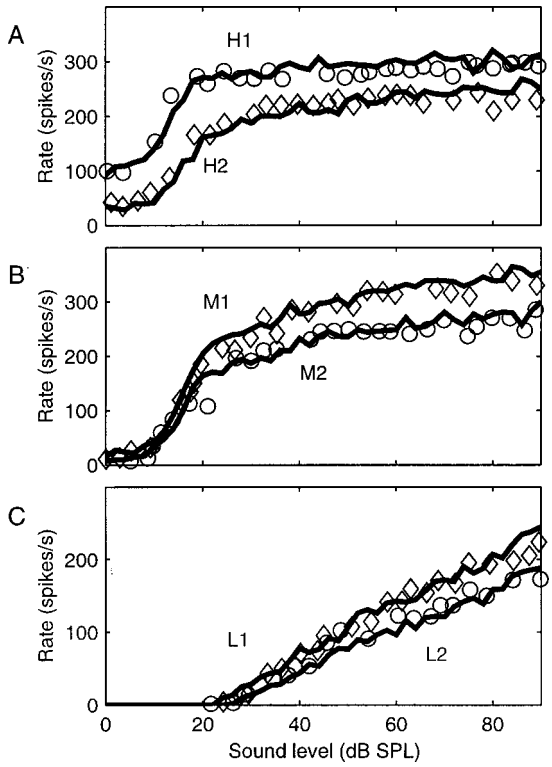


FIG. 3. AN rate-intensity responses for different fiber types. Unconnected symbols are animal data (Winter *et al.*, 1990). Solid lines are responses of the model. All fits employ the same DRNL parameters (''AN set''; Table III). Variation in synapse parameter values is shown explicitly in Table II.

[Yates *et al.*, 1990; Figs. 2(A) and 3(A)]. Both fibers are from the same guinea pig. In Fig. 4(A), a BF stimulus results in a typical HSR saturating function with a narrow dynamic range. However, when stimuli at other frequencies are used, the RI function for the same fiber can be either sloping saturation (19 kHz) or saturating (12 kHz). Yates *et al.* (1990) propose that this is due in part to the BM response, which is nonlinear at BF but linear away from BF. The DRNL filter used in this evaluation was designed to simulate this effect. Figure 4(C) shows the response of an MSR fiber from the same animal with a similar BF ($\sim$16 kHz). Figures 4(B) and (D) show the model RI response to BF and off-BF tones. Both fiber responses use the same DRNL parameters (see Table III, ''AN set'') that were chosen specifically for these data. The difference between the two figures arises from changes to calcium control parameters $G_{\text{Ca}}^{\max}$ and $[\text{Ca}^{2+}]_{\text{thr}}$. These values are given in Table II [columns labeled HSR and MSR for Figs. 4(B) and (D), respectively]. The good agreement between the model and the data indicates that it is possible to simulate important characteristics of different fiber types in terms of differences in synapse parameters.

3. The influence of the transmitter release function on fiber type

Figure 5(A) shows how model RI functions vary when the maximum calcium conductance parameter, $G_{\text{Ca}}^{\max}$, is varied. Other parameters are the same as those for the model shown in Fig. 4(A). As conductance increases, the function changes from ''straight'' to ''sloping saturation,'' and finally

TABLE II. IHC synapse calcium channels and transmitter release [Eqs. (4–11)].

z, scalar $(s*[Ca^{2+}]^3)^{-1}$	20×10^{31}
E_{Ca}, reversal potential (V)	0.066
β_{Ca}	400
γ_{Ca}	130
τ_m, calcium current time constant (s)	1×10^{-4}
τ_{Ca}, calcium diffusion time constants (s)	1×10^{-4}
y, replenishment rate (s^{-1})	10
l, loss rate (s^{-1})	2580
x, reprocessing rate (s^{-1})	66.3
r, recovery rate (s^{-1})	6580

Fiber ID	HSR	MSR	H_1	H_2	M_1	M_2	L_1	L_2
G_{Ca}^{max}, max. Ca^{2+} conductance (nS)	8	4.5	7	4.5	4	4.25	2.75	2.75
$[Ca^{2+}]_{thr}(\times10^{-11})$, threshold Ca^{2+} conc.	4.48	3.2	2	0	2	2.5	4	4.2
M, max. free transmitter quanta	10	10	10	8	13	9	8	6

to ''saturating.'' At the same time, the threshold drops, the spontaneous rate increases, and the synapse saturates at lower stimulus intensities. Calcium conductance alone seems to determine, at least qualitatively, the full range of RI function types, spontaneous rates, and thresholds described in the literature.

The effect of varying $[Ca^{2+}]_{thr}$, for two different values of G_{Ca}^{max}, corresponding approximately to low- and high-spontaneous rate responses, is shown in Fig. 5(B). $[Ca^{2+}]_{thr}$ imposes a threshold calcium concentration for transmitter release [Eq. (8)] and affects the response mainly at low intensities. Therefore, it is not possible to model the observed range of fiber characteristics by changing $[Ca^{2+}]_{thr}$ alone. However, we have used $[Ca^{2+}]_{thr}$ in the preceding figures to improve the exact fit to the data beyond that possible using G_{Ca}^{max} alone (particularly in defining the precise threshold of LSR fibers). The effect of varying the maximum number of vesicles available for immediate release, M, for two different values of G_{Ca}^{max} is shown in Fig. 5(C). The only effect of M is to scale the release rate linearly across the entire dynamic range. Therefore, M alone cannot account for differences in fiber type. A larger value of M can be helpful in fitting MSR and LSR responses, as it increases the firing rate, without driving the synapse into saturation. M is 13 or less in all cases.

B. Phase locking

The ability of the transduction apparatus to follow, cycle-by-cycle, a stimulus waveform, in all mammals, falls

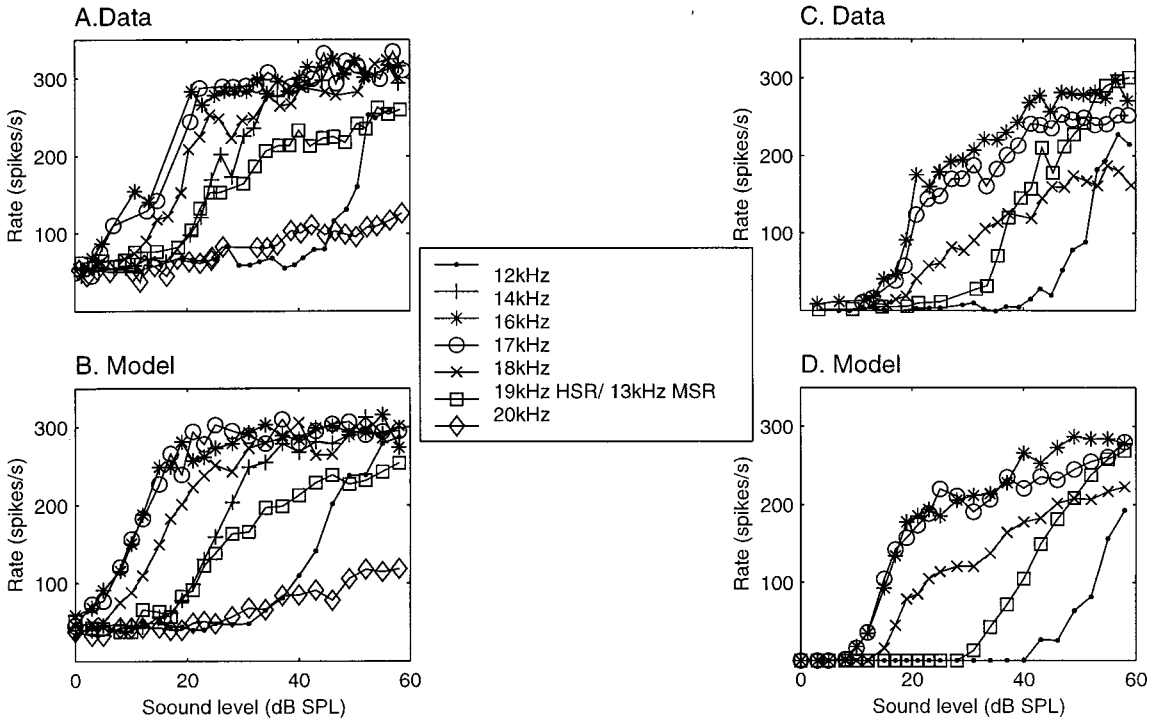


FIG. 4. Rate-intensity responses of AN fibers both at BF and off-BF. (A) HSR fiber from Yates *et al.* (1990). (B) Response of model using the ''AN set'' DRNL parameters (Table III) and ''HSR'' synapse parameters ($G_{Ca}^{max}=8\times10^{-9}$,$[Ca^{2+}]_{thr}=4.48\times10^{-11}$,$M=10$). (C) MSR fiber from Yates *et al.* (1990). (D) Response of model using the ''AN set'' DRNL parameters and ''MSR'' synapse parameters ($G_{Ca}^{max}=4.5\times10^{-9}$,$[Ca^{2+}]_{thr}=3.2\times10^{-11}$,$M=10$).

TABLE III. Basilar-membrane filtering (DRNL) parameters [Eq. (12)].

	IHC set	AN set
Center frequency of nonlinear path (Hz)	17 300	16 700
Bandwidth (Hz) of nonlinear path (Hz)	1 200	3 730
Gamma-tone filter cascade of nonlinear path	3	4
Low-pass filter cascade of nonlinear path	3	2
a, compression parameter	3 000	18 000
b, compression parameter	0.06	7.8×10^{-3}
ν, compression parameter	0.25	0.16
Center frequency linear path (Hz)	13 700	12 900
Bandwidth of linear path (Hz)	1 400	800
Gamma-tone filter cascade of linear path	2	2
Low-pass filter cascade of nonlinear path	4	3
Linear path gain	720	780

off with frequency. In the guinea pig, the decay is rapid above 1–2 kHz. Here, we compare the model response to the RP and AN observations of Palmer and Russell (1986).

1. ac/dc ratios of the receptor potential

Phase locking in the hair cell can be measured by the ratio of the peak-to-trough magnitude of the intracellular potential, to its dc displacement from rest. Figure 6(A) compares the data of Palmer and Russell (various symbols) with the response of the model (solid line). Palmer and Russell

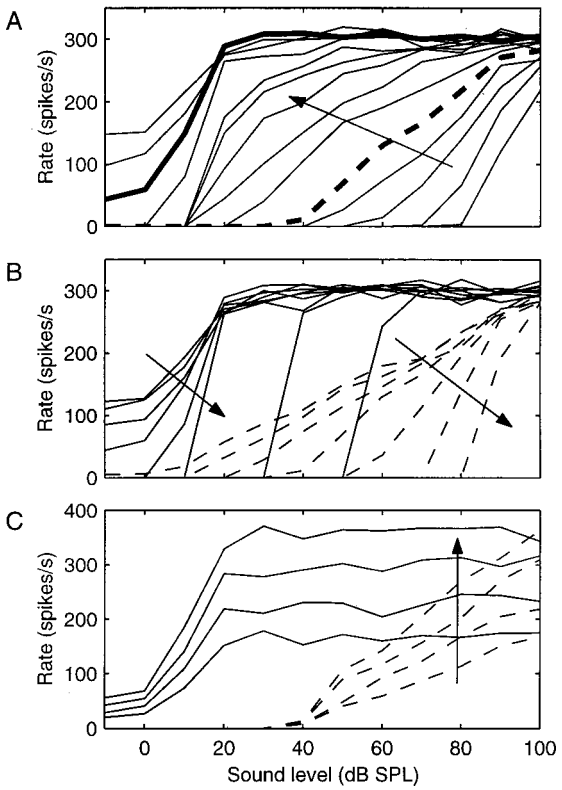


FIG. 5. Effect of varying synapse parameters on rate-intensity functions, starting from the ''HSR'' synapse values (Table II). The ''AN set'' (Table III) DRNL parameters were used. (A) The effect of increasing G_{Ca}^{max}, from 1.5 to 10 nS in the direction of the arrow. Thick solid line is for $G_{Ca}^{max}=8$ nS and thick dashed line is for $G_{Ca}^{max}=2.5$ nS. (B) The effect of increasing $[Ca^{2+}]_{thr}$ from zero to 18×10^{-11} in the direction of the arrows, for two different values of G_{Ca}^{max} (2.5 nS for dotted lines and 8 nS for solid lines). (C) The effect of increasing M from 5 to 13 in the direction of the arrows, for the same two values of G_{Ca}^{max} as (B) (for $[Ca^{2+}]_{thr}=4.48\times10^{-11}$).

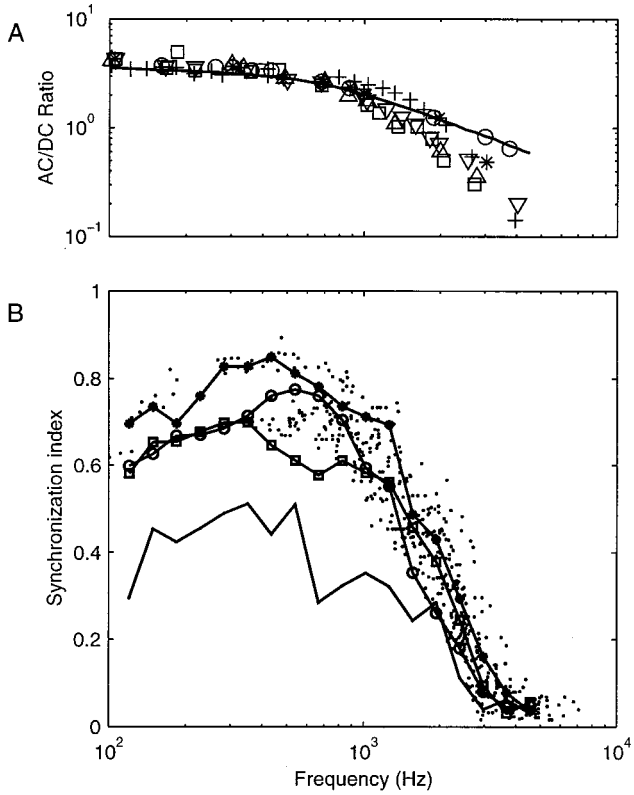


FIG. 6. Phase-locking. (A) Receptor potential ac/dc ratio. Unconnected symbols are several datasets from the guinea pig data (Palmer and Russell, 1986); the solid line is the response of the model. The DRNL parameters used are given in Table III, ''IHC set.'' (B) The synchronization index at the AN. Unconnected dots are guinea-pig data and the solid lines are model responses at four signal levels (——10 dB; □ 0 dB, ● 20 dB, ○ 40 dB SL). The model uses the ''HSR'' synapse values (Table II) and the ''AN set'' (Table III) DRNL parameters.

(1986) used low-frequency pure-tone stimuli (<5 kHz) at 80 dB SPL, and recorded from high-BF (>15 kHz) IHCs. It was necessary to change the low-frequency cutoff of the model middle-ear response down to 500 Hz, in order for the model to produce large enough output at this level. Such an action is supported experimentally. Even a small hole in the bulla, as used in these experiments to equalize middle-ear pressure and allow monitoring of the compound action potential, causes a strong increase in response at low frequencies (Ruggero *et al.*, 1990). The value of the membrane capacitance of the cell, C_m, was set to 6 pF, to produce the correct ac/dc ratios. This is close to the experimental findings of Russell and Sellick (1978), who found whole-cell capacitances in the range 8–16 pF, and Kros and Crawford (1990), who estimated values of 6–13 pF. The model gives a reasonable fit, although the slope is shallow compared with most of the data sets. The model used the DRNL parameter set shown in Table III under ''IHC set,'' and the RP stage parameters shown in Table I.

2. Synchronization index in the AN

The synchronization index (SI; Johnson, 1980), a measure of phase locking, is given in Fig. 6(B) for the data of Palmer and Russell (1986) and the model of Fig. 4(B). The model used the DRNL parameter set shown in Table III under ''AN set,'' the RP stage parameters shown in Table I, and

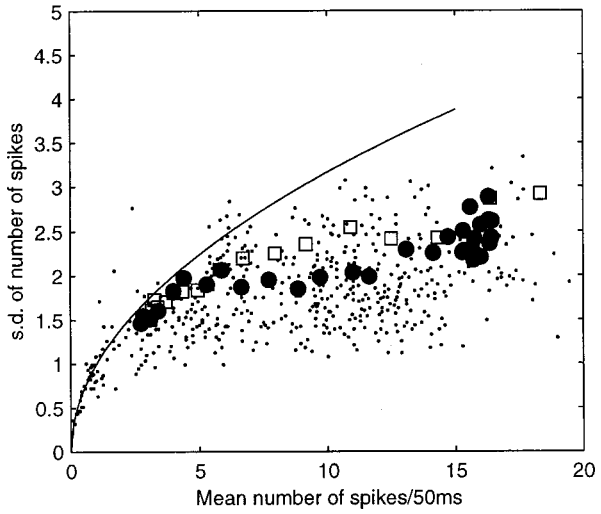


FIG. 7. The relationship of mean to variance in AN firing rate for 50-ms tone pips as measured by Winter and Palmer (1991). Small dots are the data of Winter and Palmer; solid circles are the behavior of the model using the "AN set" DRNL parameters (Table III) and "MSR" synapse parameters (Table II). Open squares are the response of this model without the explicit absolute and relative refractoriness of Sec. III C.

"HSR" synapse parameters in Table II. The time constants for calcium influx, τ_m, and accumulation, τ_{Ca}, were both set to 0.1 ms. These values produced a good fit to the data, and are consistent with the results of Hudspeth and Lewis (1988). The animal phase-locking measurements were made in fibers having BFs below 9 kHz. We used a high BF in these simulations and, as a consequence, the sound level required is considerably higher. However, with small exceptions, there is a good fit between the model results for 20-dB SL. The thresholds vary from 85–110 dB SPL. At these frequencies, the output from the DRNL is linear, so the comparison is valid despite the high sound levels. The effect of sound level on SI is not shown by any animal data here. Qualitatively, Palmer and Russell noted that SI rises with level, is often detectable below threshold, and reaches a maximum within 20 dB above threshold. Above that, they found that SI sometimes decreases slightly. Figure 6(B) shows the model response from −10 to 40-dB SL. The behavior of the model conforms to Palmer and Russell's findings. The changes in SI with sound level in the model are a consequence of the fact that random spontaneous activity compromises phase locking at low signal levels. At very high levels, the signal is above threshold for a greater fraction of its period, thus reducing the SI.

C. Spike statistics

1. Mean-to-variance of firing rate

A number of studies have been made of the relationship between the mean firing rate and its variance (e.g., Teich and Khanna, 1985; Young and Barta, 1986; Winter and Palmer, 1991). At higher firing rates the variance is less than the mean, indicating a departure from the prediction of a Poisson model of spiking statistics. The difference is often accounted for by introducing a simple refractory stage (e.g., Young and Barta, 1986). Figure 7 shows the mean-variance relationship

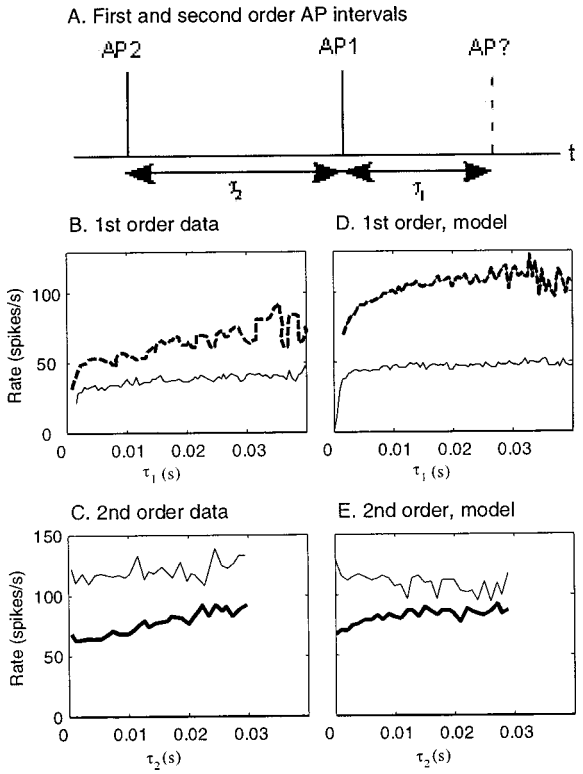


FIG. 8. AN discharge history effects. (A) Cartoon of APs showing first- and second-order intervals. Solid bars (AP1 and AP2) are spikes that have already occurred and the dashed bar is a possible spike (AP?) at the instant τ_1 after the immediately preceding spike AP1. (B) First-order hazard functions derived from spontaneous APs of cat from Gaumond *et al.* (1982, thin solid line) and Grey (1967, thick dashed line). τ_1 is the time interval between two APs. (C) Second-order hazard functions from Gaumond *et al.* τ_2 is the time interval between the preceding spike (AP1) and the spike before it (AP2). The thick line shows the hazard function based on short values of τ_1 between 1 and 4 ms. The thin line is based on longer intervals, τ_1 between 20 and 40 ms. (D) First-order hazard function for spontaneous activity of the model. The thin solid line uses the "AN set" DRNL parameters (Table III) and "HSR" synapse parameters (Table II). The thick dashed line uses the modified model (see the text). (E) Second-order hazard functions for the modified model fiber of (D), for comparison with (C).

measured by Winter and Palmer (1991) in the guinea pig (small dots). A clear departure from the Poisson prediction (solid line) is evident. The large solid circles represent the behavior of the model for the MSR fiber (the HSR model could not be used for measurements at spike rates below its SR), applied to stimulus paradigm described by Winter and Palmer. The model agrees well with the data, considering that no attempt was made to tune the model for this demonstration. The output of the model is also plotted with the AN refractory module removed (open squares). There is still considerable departure from the Poisson prediction with the synapse alone.

2. Discharge history effects

During steady-state activity, the instantaneous rate of firing of AN fibers is depressed immediately after an AP (first-order effects) and further depressed if two APs follow in rapid succession (second-order effects). Figure 8(A) shows this. First-order effects can be quantified, using hazard functions[2] that estimate the probability of an AP [AP? in Fig. 8(A)] as a function of the interval, τ_1, since the last spike

AP1. The animal hazard functions of spontaneous activity in Fig. 8(B) (Grey, 1967; thick dashed line; Gaumond *et al.*, 1982; thin solid line) show a depression in firing rate following short interspike intervals (small τ_1) that recovers over a 30-ms period. However, as shown in Fig. 8(B), the magnitude of the effect varies between observations. Second-order discharge history effects can be quantified as the probability of a spike, AP?, as a function of the interval, τ_2, between the preceding spike, AP1, and the one before that, AP2 [see Fig. 8(A)]. This is normally calculated as an average across a given range of the interval, τ_1. Figure 8(C) shows two hazard series based on a long train of spontaneous APs (Gaumond *et al.*, 1982), where each series is expressed as a function of the preceding interspike interval, τ_2. The thick line deals with very short first-order intervals ($1<\tau_1<4$ ms), while the thin line deals with longer intervals ($20<\tau_1<40$ ms). The firing rate is depressed most when short first-order interspike intervals (τ_1, thick line) are combined with short preceding second-order intervals (low values of τ_2).

Discharge-history effects can also be seen in the working of the model where it is caused by transmitter depletion. A similar analysis of a train of spontaneous APs from the model is shown in Figs. 8(D) and (E). In the standard HSR AN model, first-order effects [Fig. 8(D), thin solid line] are very small, except for the first few milliseconds, which are dominated by AN refractory effects [Eq. (13)]. The first-order effects for longer intervals are nonetheless consistent with Gaumond *et al.*'s data in Fig. 8(B) (thin solid line). Second-order effects (not shown) are similarly small. The thick dashed line in Fig. 8(D) shows the first-order effects with the model parameters modified to increase discharge history effects. This model shows a trend more similar to that of Grey's data [thick dashed line in Fig. 8(B)]. The spontaneous rate of the model fiber was increased by setting $G_{\mathrm{Ca}}^{\max}$ to 11 nS in combination with a reduction in the maximum number of available transmitter vesicles, M, to 5. This latter value is important because the loss of a single vesicle from the immediate store, $q(t)$, represents a substantial fall in the probability of a succeeding action potential when $q(t)$ is already small. Figure 8(E) shows the second-order effects of the modified model. The model results are comparable with the animal data.

V. DISCUSSION

The model is intended to act as one module in a much larger system. Its success as a computational tool, therefore, depends primarily on the range of IHC phenomena that can be simulated. It has been shown that the model successfully reproduces the BF-tone RI functions for all three fiber types (HSR, MSR, LSR), and most of the responses to stimuli away from BF. We note that Jackson and Relkin (1998) have found that AN fiber saturated rates drop with increasing stimulus frequency above BF. The model has not reproduced this effect. It successfully reproduces correct phase-locking characteristics at the level of both the RP and the AN; the departure of AN spiking statistics from the Poisson prediction, and first- and second-order discharge-history effects. There are a number of other phenomena that have not been tested here. Adaptation has been the focus of earlier studies (Meddis, 1986, 1988). Differences of adaptation with fiber type will be a topic of future study. The response to broadband signals and, notably, two-tone suppression (2TS), is thought to have its origin in the BM. An evaluation of the DRNL filter can be found elsewhere (Meddis *et al.*, 2001). Two-tone suppression will be reported elsewhere.

The varied scope of modeling in the auditory periphery makes meaningful comparisons between models difficult. This is especially true of older models. Perhaps the closest of the recent complete peripheral models are those of Zhang *et al.* (2001) and Robert and Erikson (1999). These achieve many of the same things for the cat as we do for the guinea pig. As computational tools, all are very effective. However, our approach to the task differs. This study concentrates on the development a detailed IHC model. Zhang *et al.* and Robert and Erikson focused on two-tone suppression while we have not. We have evaluated the IHC model only for a high-BF site (~18 kHz). The choice of high BFs was decided by the excellent availability of data for the guinea pig. At low BFs, data are less complete, and the contributions of mechanical and biophysical nonlinearities are not as well understood. Nevertheless, there is no strong reason to suppose that IHCs differ substantially in their response along the partition. A complete filterbank will be described elsewhere. Our preferred strategy was to simulate, whenever known, the underlying component physiological processes in the hope that this would lead to better overall performance. A notable success in this respect is the spiking statistics arising from the stochasticity of the synapse.

Our model is most similar to the IHC models of Lopez-Poveda *et al.* (1998) and Schoonhoven *et al.* (1997). They both employed *ad hoc* transmitter release control functions, while, in this study, we have taken a further step, and introduced a model of calcium control of transmitter release. The differences in the model between HSR, MSR, and LSR AN fibers are attributed largely to differences in the presynaptic maximum calcium conductance ($G_{\mathrm{Ca}}^{\max}$). No other parameter had this effect. The characteristics of the calcium current are based on physiological observations, and it seems reasonable that transmitter release will be affected if the number of calcium channels vary. However, the model can only establish the feasibility of this mechanism. The calcium model is very simple and it is possible that other mechanisms not known or not included in the model (e.g., calcium buffering) may be responsible for fiber-type differences. For example, it has been suggested that different calcium channel types may underlie the differences between fiber types (Zhang *et al.*, 1999). Mechanisms might also be postsynaptic (e.g., Geisler, 1981). Very recently, Puel (2001) has demonstrated that efferents synapsing onto the dendrites of the AN regulate SR, thresholds, and dynamic range of AN fibers via tonic inhibition.

The use of two cubed functions in the transmitter release function produces a ninth-order function of sound level [Eq. (4) and Eq. (8)]. This may appear to contradict some experimental data, showing that dc RP and AN firing obeys square-law behavior near to threshold (Dallos, 1985; Müller *et al.*,

1991). However, the functions employed for the calcium processes are themselves based on experimental data (Augustine *et al.*, 1985a, b). These functions do not hamper the model for the RP or AN data shown here, it remains an issue to be resolved.

In a modeling study, Kidd and Weiss (1990) found that calcium dynamics is likely to impact on IHC phase locking. Calcium dynamics plays a role in our model, but its effect on phase locking is relatively minor. Other factors, including the receptor potential and delays in the circulation of the transmitter material, also have an effect. However, our model does not allow us to draw any strong conclusion about the relative importance of calcium in phase locking. Although the AN output of the model fits the phase-locking data excellently, the fit to the RP data is not perfect. This is probably because observed voltage-dependent potassium currents (Kros and Crawford, 1990) are modeled as a passive conductance. Further research is required before we can effectively model the contribution of each process.

A new feature of our model is the location of stochasticity at the synapse. In most other models, the auditory-nerve stage is itself treated as stochastic. However, studies of isolated spiral ganglion cells (Lin, 1997; Santos-Saachi, 1993) show little evidence of stochasticity. Likewise, the required degree of stochasticity is not observed in RP recordings. In contrast to most synapses, auditory-nerve synapses have a relatively small vesicle supply (Liberman, 1980), so the stochastic release of vesicles is probably unavoidable. Employing quantal vesicle release in the model means that the value of M, the maximum number of vesicles available for release, is an important determinant of synaptic response. Small values of M, typically less than 13, are essential to give appropriate results. When modeling Gaumond *et al.*'s data (1982), it was necessary to restrict the value of M to 5 quanta in order to reproduce his pronounced second-order discharge history effects. The small number of quanta typically required for effective functioning of the model amounts, in effect, to a prediction requiring physiological confirmation when techniques for investigating individual synapses become sophisticated enough to address the issue. The synapse in the model also accounts for the majority of the observed mean-to-variance relationship. This is interesting, as it is usually attributed to the primary auditory neurons. The introduction of the stochasticity in the synapse substantially reduces the contribution of the AN refractoriness. We found, however, that the sub-1-ms effects could only be explained if the model included the 0.75-ms dead time and the additional short-term relative refractory effect.

[1]The modules employed are available in the AMS (Auditory Model Simulator) application. Both DSAM and AMS are available from http:\\www.essex.ac.uk\psychology\hearinglab\dsam

[2]This is calculated from an interspike-interval histogram. Expressed as a firing rate, it is the ratio of the number of intervals τ_1 over the number of intervals $\geqslant \tau_1$, all divided by the bin width (0.5 ms).

Augustine, G. J., Charlton, M. P., and Smith, S. J. (**1985a**). ''Calcium entry into voltage-clamped pre-synaptic terminals of squid,'' J. Physiol. (London) **367**, 143–162.

Augustine, G. J., Charlton, M. P., and Smith, S. J. (**1985b**). ''Calcium entry and transmitter release at voltage-clamped nerve terminals of squid,'' J. Physiol. (London) **367**, 163–181.

Carney, L. H. (**1993**). ''A model for the responses of low-frequency auditory-nerve fibers in cat,'' J. Acoust. Soc. Am. **93**, 401–417.

Corey, D. P., and Hudspeth, A. J. (**1983**). ''Kinetics of the receptor current in the bullfrog saccular hair-cells,'' J. Neurosci. **3**, 962–976.

Dallos, P. (**1985**). ''Response characteristics of mammalian cochlear hair cells,'' J. Neurosci. **5**, 1591–1608.

Davis, H. (**1965**). ''A model for transducer action in the cochlea,'' Cold Spring Harbor Symp. Quant. Biol. **30**, 181–189.

Eggermont, J. J. (**1985**). ''Peripheral auditory adaptation and fatigue: A model oriented review,'' Hear. Res. **18**, 57–71.

Gaumond, R. P., Molnar, C. E., and Kim, D. O. (**1982**). ''Stimulus and recovery dependence of cat cochlear nerve fiber spike discharge probability,'' J. Neurophysiol. **48**, 856–873.

Geisler, C.D. (**1981**). ''A model for discharge patterns of primary auditory-nerve fibers,'' Brain Research **212** 198–201.

Grey, P. R. (**1967**). ''Analyses of spike activity of single neurons,'' Biophys. J. **7**, 759–777.

Hewitt, M. J., and Meddis, R. (**1991**). ''An evaluation of eight computational models of mammalian inner hair-cell function,'' J. Acoust. Soc. Am. **90**, 904–917.

Hudspeth, A. J., and Lewis, R. S. (**1988**). ''Kinetic analysis of voltage- and ion-dependent conductances in saccular hair cells of the bull-frog, *rana catesbeneiana*,'' J. Physiol. (London) **400**, 237–274.

Irino, T., and Patterson, R. D. (**2001**). ''A compressive gammachirp auditory filter for both physiological and psychophysical data,'' J. Acoust. Soc. Am. **109**, 2008–2022.

Jackson, B. S., and Relkin, E. M. (**1998**). ''A frequency-dependent saturation evident in rate-intensity functions of the chinchilla auditory nerve,'' Hear. Res. **126**, 75–83.

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.

Johnston, D., and Wu, S. M. (**1995**). *Foundations of Cellular Neurophysiology* (MIT University Press, Cambridge, MA).

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 University Press, Cambridge, MA).

Kidd, R. C., and Weiss, T. F. (**1990**). ''Mechanism that degrade timing information in the cochlea,'' Hear. Res. **49**, 181–208.

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. (**1980**). ''Morphological differences amoung radial afferent fibers in the cat cochlea: An electron-microscopic study of serial sections,'' Hear. Res. **3**, 45–63.

Lin, X. (**1997**). ''Action potentials and underlying voltage-dependent currents studied in cultured spiral ganglion neurons of the postnatal gerbil,'' Hear. Res. **108**, 157–179.

Lopez-Poveda, E. A., O'Mard, L. P., and Meddis R. (**1998**). ''A revised computational inner hair cell model,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 102–108.

Meddis, R. (**1988**). ''Simulation of auditory-neural transduction: Further studies,'' J. Acoust. Soc. Am. **83**, 1056–1063.

Meddis, R. (**1986**). ''Simulation of mechanical to neural transduction in the auditory receptor,'' J. Acoust. Soc. Am. **79**, 702–711.

Meddis, R., and Hewitt, M. J. (**1991**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., Hewitt, M., and Shackleton, T. (**1990**). ''Implementation details of a computational model of the inner hair-cell/auditory–auditory nerve synapse,'' J. Acoust. Soc. Am. **87**, 1813–1816.

Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). ''A computational algorithm for computing nonlinear auditory frequency selectivity,'' J. Acoust. Soc. Am. **109**, 2852–2861.

Mountain D. C., and Hubbard, A. E. (**1996**). ''Computational analysis of hair cell and auditory nerve processes,'' in *Auditory Computation* edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer, New York).

Müller, M., Robertson D., and Yates, G. K. (**1991**). ''Rate-versus-level functions of primary auditory-nerve fibers—evidence for square law behavior of all fiber categories in the guinea-pig,'' Hearing Res. **55**, 50-56.

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**, 1556–1565.
Palmer, A. R., and Russel, 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.
Patterson, R. D., Nimmo-Smith, I., and Holdsworth, J., and Rice, P. (**1988**). ''Spiral vos final report, Part A: The auditory filterbank,'' Cambridge Electronic Design, Contract Rep. (Apu 2341).
Patuzzi, R., and Sellick, P. M. (**1983**). ''A comparison between basilar membrane and inner hair cell receptor potential input–output functions in the guinea pig cochlea,'' J. Acoust. Soc. Am. **74**, 1734–1741.
Puel, J. L. (**2001**). ''*In vivo* studies of lateral olivocochlear efferents,'' Assoc. Res. Otolaryngol. Abs: #492, 138.
Rattay, F., Gebeshuber, I. C., and Gitter, A. H. (**1998**). ''The mammalian auditory hair cell: A simple electric circuit model,'' J. Acoust. Soc. Am. **103**, 1558–1565.
Robert, A., and Eriksson, J. L. (**1999**). ''A composite model of the auditory periphery for simulating responses to complex sounds,'' J. Acoust. Soc. Am. **106**, 1852–1864.
Ruggero, M. A., Rich, N. C., Robles, L., and Shivapuja, B. G. (**1990**). ''Middle-ear response in the chinchilla and its relationship to mechanics at the base of the cochlea,'' J. Acoust. Soc. Am. **87**, 1612–1629.
Russell, I. J., and Sellick, P. M. (**1978**). ''Intracellular studies of hair cells in the mammalian cochlea,'' J. Physiol. (London) **284**, 261–290.
Russell, I. J., Cody, A. R., and Richardson, G. P. (**1986**). ''The responses of inner and outer haircells in the basal turn of the guinea-pig cochlea and in the mouse cochlea grown *in vitro*,'' Hear. Res. **22**, 199–216.
Santos-Sacchi, J. (**1993**). ''Voltage-dependant ionic conductances of type I spiral ganglion cells from the guinea pig inner ear,'' J. Neurosci. **13**, 3599–3611.
Schoonhoven, R., Prijs, V. F., and Frijns, J. H. M. (**1997**). ''Transmitter release in inner hair cell synapses: A model analysis of spontaneous and driven properties of cochlear nerve fibers,'' Hear. Res. **113**, 247–260.
Schroeder, M. R., and Hall, J. L. (**1974**). ''Model for mechanical to neural transduction in the auditory receptor,'' J. Acoust. Soc. Am. **55**, 1055–1060.
Shamma, S. A., Chadwick, R. S., Wilbur, W. J., Morrish, K. A., and Rinzel, J. (**1986**). ''A biophysical model of the cochlear processing: Intensity dependence of pure tone responses,'' J. Acoust. Soc. Am. **80**, 133–145.
Siegel, J. H. (**1992**). ''Spontaneous synaptic potentials from afferent terminals in the guinea pig cochlea,'' Hear. Res. **59**, 85–92.
Smith, R. L., and Brachman, M. L. (**1982**). ''Adaptation in auditory-nerve fibers: A revised model,'' Biol. Cybern. **44**, 107–120.
Teich, M. C., and Khanna, S. M. (**1985**). ''Pulse-number distribution for neural spike train in the cat's auditory nerve,'' J. Acoust. Soc. Am. **77**, 1110–1128.
Weiss, T. F. (**1966**). ''A model of the peripheral auditory system,'' Kybernetik **3**, 153–175.
Winter, I. M., Robertson, D., and Yates, G. K. (**1990**). ''Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibers,'' Hear. Res. **45**, 191–202.
Winter, I. M., and Palmer, A. R. (**1991**). ''Intensity coding in low-frequency auditory nerve fibers of the guinea-pig'', J. Acoust. Soc. Am. **90**, 1958–1967.
Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). ''Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range,'' Hear. Res. **45**, 203–220.
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.
Zhang, S. Y., Robertson, D., Yates, G., and Everett, A. (**1999**). ''Role of L-type Ca2+ channels in transmitter release from mammalian inner hair cells I. Gross sound-evoked potentials,'' J. Neurophysiol. **82**, 3307–3315.
Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). ''A phenomenological model for the responses of auditory nerve fibers. I. Nonlinear tuning with compression and suppression,'' J. Acoust. Soc. Am. **109**, 648–670.

Vibration measurement of the tympanic membrane of guinea pig temporal bones using time-averaged speckle pattern interferometry

Hiroshi Wada,[a] Masayoshi Ando, Masataka Takeuchi, Hironori Sugawara, and Takuji Koike
Department of Mechanical Engineering, Tohoku University, Sendai 980-8579, Japan

Toshimitsu Kobayashi and Koji Hozawa
Department of Otolaryngology, Tohoku University School of Medicine, Sendai 980-8574, Japan

Takashi Gemma
Department of Optical Design Headquarters, Nikon Corporation, Nishiohi 1-6-3, Shinagawa-ku, Tokyo 141-8601, Japan

Makoto Nara
Department of Design, Nikon Technologies, Incorporated, Futaba 1-3-25, Shinagawa-ku, Tokyo 142-0043, Japan

(Received 6 June 2001; revised 12 February 2002; accepted 13 February 2002)

''Time-averaged holography'' and ''holographic interferometry'' enable recording of the complete vibration pattern of a surface within several seconds. The results appear in the form of fringes. Vibration amplitudes smaller than 100 nm are not readily measurable by these techniques, because such small amplitudes produce variations in gray level, but not fringes. In practice, to obtain clear fringes in these measurements, stimulus sound pressures higher than 100 dB SPL must be used. The phase of motion is also not obtainable from such fringe techniques. In this study, a sinusoidal phase modulation technique is described, which allows detection of both small amplitudes of motion and their phase from time-averaged speckle pattern interferometry. In this technique, the laser injection current is modulated and digital image processing is used to analyze the measured patterns. When the sound-pressure level of stimuli is between 70 and 85 dB SPL, this system is applied to measure the vibratory response of the tympanic membrane (TM) of guinea pig temporal bones at frequencies up to 4 kHz where complicated vibration modes are observed. The effect of the bulla on TM displacements is also quantified. Results indicate that this system is capable of measuring the nanometer displacements of the TM, produced by stimuli of 70 dB SPL. [DOI: 10.1121/1.1467671]

PACS numbers: 43.64.Ha, 43.64.Yp [BLM]

I. INTRODUCTION

Several methods have been developed to measure small displacement amplitude of tympanic membrane (TM) vibrations. One of them, tympanometry (ANSI, 1987; Lilly, 1984; Wada and Kobayashi, 1990), is useful for aiding in the differential diagnosis of middle-ear diseases. However, this method is a measure of the average motion of the entire TM and does not distinguish the differential motion of different locations on the membrane. By contrast, a commercially available laser Doppler velocimeter (LDV) (Buunen and Vlaming, 1981; Goode *et al.*, 1993; Nishihara *et al.*, 1993; Vlaming and Feenstra, 1986a, b) is a sensitive, noncontacting device capable of measuring the displacement amplitude of TM vibrations down to 1 nm. However, an LDV is generally not used to measure the vibration pattern of the entire TM, except in cases where a scanning LDV system is employed (Konradsson *et al.*, 1987; Goode *et al.*, 1996; Ball *et al.*, 1997). Heterodyne interferometry is also a very sensitive measurement technique. Decraemer, Khanna, and Funnell (1989, 1991) applied this technique to measure precisely the amplitude and phase of TM and malleus vibrations in cats at several points, and examined the vibration mode of the manubrium.

Time-averaged holography and holographic interferometry enable us to record the complete vibration pattern of a surface within several seconds (Saldner, 1996). Results appear in the form of fringes, in other words, alternating bright and dark lines. Each line represents an isoamplitude contour, and it is possible to precisely determine the absolute amplitude corresponding to each of the contour lines. Khanna and Tonndorf (1972) and Tonndorf and Khanna (1972) applied time-averaged holography to analyze the dynamic behavior of the TM in cats and human cadavers, and showed that the TM did not vibrate like a stiff plate as observed by von Bekesy (1949). Using holographic interferometry, Dancer *et al.* (1975) investigated TM displacement in the guinea pig subjected to acoustic impulses. Vibration amplitude smaller than 100 nm, however, is not obtainable by these techniques, because smaller amplitude does not result in fringes, but only in variations in gray levels, which are difficult to quantify.

[a]Electronic mail: wada@cc.mech.tohoku.ac.jp

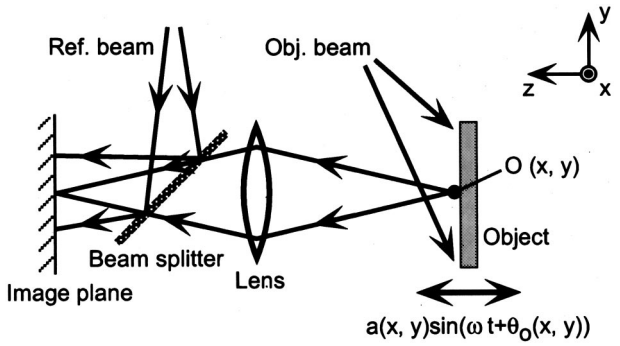


FIG. 1. Principle of interferometry. A laser beam is split into two beams, one of which is reflected from the object, which vibrates sinusoidally. The other beam is reflected in the beam splitter and is combined in line with the object beam. The incoming light intensity is accumulated at the image plane.

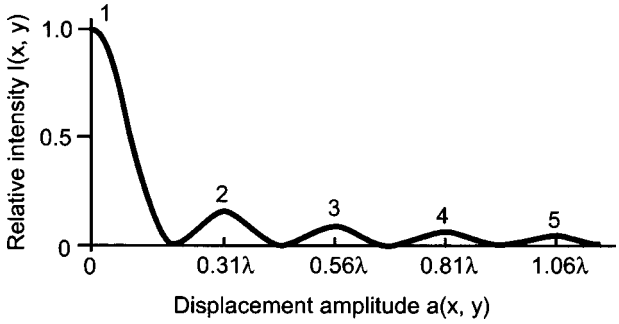


FIG. 2. Relationship between the intensity and the displacement amplitude of the object, which is given by Eq. (1). Each peak is numbered.

Therefore, to obtain clear fringes, they supplied a sound pressure higher than 100 dB SPL in their measurements. Phase information is also not obtainable by their techniques.

For the purpose of detecting smaller amplitude and obtaining phase information, the phase modulation technique was developed (Lokberg and Hogmonen, 1976; Johns and Wykes, 1989; Ellingsrud and Rosvold, 1992; Petzing and Tyrer, 1998). By use of time-averaged speckle pattern interferometry (SPI) combined with the sinusoidal phase modulation (SPM), it is said that the amplitude detection limits are reduced from 100 to 2 nm. Furthermore, the clear phase information is obtainable. Using this technique, the TM frequency response at the malleus head in a human temporal bone was measured by Hogmonen and Lokberg (1977), and that at the malleus manubrium in humans was examined *in vivo* by Lokberg *et al.* (1980). They also presented a sketch of the TM dynamic behavior when $f=2.4$ kHz and $P=90$ dB SPL. However, the vibration modes of the TM, except for those mentioned above, have not yet been displayed.

Owing to recent developments in electronics, the wavelength of a laser beam can be easily modulated by a direct modulation of the laser injection current. Therefore, in this study, using time-averaged electronic SPI, the SPM technique by a direct modulation of laser injection current, and the digital image processing technique, an attempt is made to develop a system for measuring a small vibration sensitively. Vibratory responses of the TM of guinea pig temporal bones are then measured by this system.

II. PRINCIPLE

First, the theory of conventional time-averaged SPI will be explained. As shown in Fig. 1, when the object $O(x,y)$ vibrates sinusoidally with the displacement amplitude $a(x,y)$, and the exposure time is sufficiently long compared with the period of the vibration frequency, the reconstructed intensity $I(x,y)$ is expressed by

$$I(x,y) \propto J_0^2((4\pi/\lambda)a(x,y)), \tag{1}$$

where J_0 is the zero-order Bessel function of the first kind, and λ is the wavelength of the laser beam. Figure 2 shows the relationship between the intensity and the displacement amplitude of the object, which is given by Eq. (1). The intensity is largest when $a(x,y)=0$, and then oscillates around 0 as $a(x,y)$ increases. The magnitude of these oscillations decreases as $a(x,y)$ increases.

Although the theory and basic principles of time-averaged SPI and the SPM technique have been well described (Johns and Wykes, 1989; Ellingsrud and Rosvold, 1992), they will be explained briefly here, to ensure understanding of how displacement distributions of objects were obtained in this study. When the semiconductor laser is supplied with a modulated injection current, the wavelength λ_1 of the laser beam incident on the object is determined by

$$\lambda_1 = \lambda + \Delta\lambda \sin(\omega t + \theta_1), \tag{2}$$

where λ is its central wavelength, $\Delta\lambda$ is the amplitude of the modulated wavelength, ω is the frequency of the modulated injection current, t is the time, and θ_1 is the phase of the modulation at time zero. If the object vibrates sinusoidally with the amplitude $a(x,y)$, the phase $\theta_0(x,y)$, and the frequency which is synchronized with that of the modulated injection current, and the exposure time is sufficiently long compared with the period of the vibration frequency, the reconstructed intensity $I(x,y)$ can be expressed by

$$I(x,y) \propto J_0^2[(4\pi/\lambda)\{a(x,y)^2 + b^2 - 2a(x,y)b\cos(\theta_0(x,y) - \theta_1)\}^{1/2}], \tag{3}$$

where $b=\Delta\lambda L/(2\lambda)$, and L is the difference in the length between the path of the object beam and that of the reference beam. Figure 3 shows the relationship between intensity and small displacement amplitude of the object, which is given by Eq. (3). As shown in Fig. 3, if the working point is set so as to meet the following equation, i.e., $b=0.1\lambda$, the linear part of Eq. (3) can be utilized. If the vibration amplitude of the object is small, the linear approximation of Eq. (3) is given by

$$I(x,y) = I_b(x,y) - ka(x,y)\cos(\theta_0(x,y) - \theta_1(x,y)), \tag{4}$$

where I_b and k are the intensity and slope at the working point, respectively.

The following is an explanation of how to obtain the unknowns, i.e., the displacement amplitude $a(x,y)$ and the phase $\theta_0(x,y)$ of the object. By shifting the phase θ_1 of the laser beam at the interval of $\pi/2$, Eq. (4) can be rewritten as

$$I_0(x,y) = I_b(x,y) - ka(x,y)\cos\theta_0(x,y), \tag{5}$$

$$I_{\pi/2}(x,y) = I_b(x,y) - ka(x,y)(x,y)\sin\theta_0(x,y), \tag{6}$$

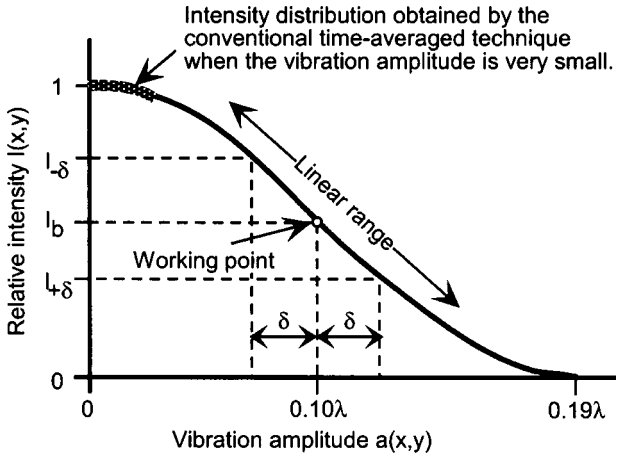


FIG. 3. Relationship between intensity and small displacement amplitude of the object, which is given by Eq. (3). When the vibration amplitude is very small, the intensity variation obtained by the conventional interferometry technique is also small. However, when the sinusoidal phase modulation technique is applied and the working point is set as shown in this figure, the linear range can be utilized and the small amplitude around this working point results in a relatively large change in intensity. In order to obtain the calibration frames $I_{+\delta(x,y)}$ and $I_{-\delta(x,y)}$, the static displacements $+\delta\,(x,\,y)$ and $-\delta\,(x,y)$ are given.

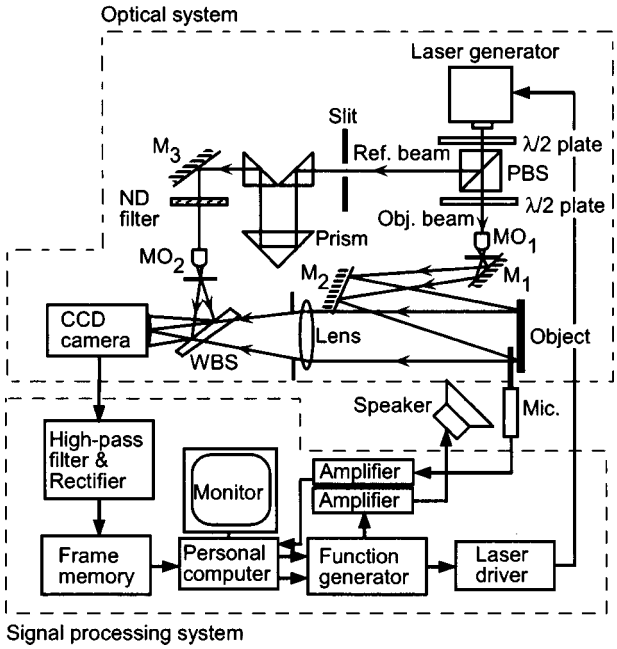


FIG. 4. Experimental setup.

$$I_{\pi}(x,y)=I_b(x,y)+ka(x,y)\cos\theta_0(x,y), \tag{7}$$

$$I_{3\pi/2}(x,y)=I_b(x,y)+ka(x,y)\sin\theta_0(x,y). \tag{8}$$

As shown in Fig. 3, in a static condition, if the calibration frames $I_{+\delta}(x,y)$ and $I_{-\delta}(x,y)$ are acquired by giving the known displacements $+\delta(x,y)$ and $-\delta(x,y)$ from the working point, Eq. (4) can be rewritten as follows:

$$I_{+\delta}(x,y)=I_b(x,y)-k\delta, \tag{9}$$

$$I_{-\delta}(x,y)=I_b(x,y)+k\delta. \tag{10}$$

From Eqs. (5)–(8), two equations are drawn in the forms

$$2ka(x,y)\cos\theta_0(x,y)=I_{\pi}(x,y)-I_0(x,y), \tag{11}$$

$$2ka(x,y)\sin\theta_0(x,y)=I_{3\pi/2}(x,y)-I_{\pi/2}(x,y). \tag{12}$$

Subtracting Eq. (10) from Eq. (9) leads to

$$2k\delta=I_{-\delta}(x,y)-I_{\delta}(x,y). \tag{13}$$

From Eqs. (11), (12), and (13), the unknown values, i.e., the displacement amplitude $a(x,y)$ and the phase $\theta_0(x,y)$ of the object, are obtained as follows:

$$a(x,y)=\delta\{(I_{3\pi/2}(x,y)-I_{\pi/2}(x,y))^2+(I_{\pi}(x,y)-I_0(x,y))^2\}^{1/2}/(I_{-\delta}(x,y)-I_{\delta}(x,y)), \tag{14}$$

$$\theta_0(x,y)=\arctan\{(I_{3\pi/2}(x,y)-I_{\pi/2}(x,y))/(I_{\pi}(x,y)-I_0(x,y))\}. \tag{15}$$

The dynamic behavior of the object is reconstructed using amplitude and phase information as follows:

$$w(x,y)=a(x,y)\cos(\theta_0(x,y)+n\pi/4), \tag{16}$$

where $w(x,y)$ is the displacement of the vibrating object and n is an integer. In order to produce the dynamic animation of the object, eight sequential instantaneous displacement images are necessary, i.e., $n=0,1,...,7$.

III. EXPERIMENTAL SETUP AND MATERIALS

An experimental setup is shown in Fig. 4. A beam emitted from the semiconductor laser generator (SDL, 7501-G1, 15 mW, $\lambda=635$ nm) is divided by the polarizing beam splitter (PBS) into the object and reference beams. The reference beam goes through the prism, and its density is controlled by the neutral density filter (ND filter). The object beam is expanded at the objective lens (MO1) with a pinhole and is diffusely reflected by the object. These two beams are combined in line with each other, and an interference speckle pattern is generated at the wedged beam splitter (WBS). This image is accumulated by use of a CCD camera (Toshiba, IK-M41MR).

In the signal-processing part, the video signal is stored in an 8-bit 256-distinct-level frame memory (Microtechnica, SIAT-MN) after the signal is high-pass filtered and rectified (Nikon, custom-made). The data are then processed by the computer (Compaq, Desk Pro XL575). The function generator (Sony Tektronix, AFG320) provides excitation signals for both the laser driver (ILX Lightwave, LDX-3525) and speaker (Foster, FE83, 8Ω, 2W), which is set close to the object and excites it in the open field. The gain of the function generator is controlled so that the value of the modulated wavelength $\Delta\lambda$ becomes 8.06×10^{-12} m. The sound-pressure level is monitored by the microphone (Etymotic, ER-7C), which is set 5 mm in distance from the object.

Sixty frames, i.e., 10 frames each of $I_{+\delta}(x,y)$, $I_{-\delta}(x,y)$, $I_0(x,y)$, $I_{\pi/2}(x,y)$, $I_{\pi}(x,y)$, and $I_{3\pi/2}(x,y)$ shown in Eqs. (14) and (15), are recorded at a rate of 30 Hz at each sound-pressure level and frequency, and six interferograms are obtained by averaging the 10 frames of each. It takes 10 s to transfer the data from the CCD camera, which has 160×120 pixels, to the frame memory and obtain the six interferograms. They are smoothed by averaging areas consisting of 5×5 consecutive pixels. After one such area is

averaged, the next area to be averaged is shifted horizontally or vertically until all such overlapping areas have been averaged. Then, the amplitude, phase, and displacement are acquired from Eqs. (14), (15), and (16).

The specifications of this measurement system are as follows: Maximum diameter of the measured object is 15 mm, maximum measurable displacement amplitude is 60 nm, which is determined by the wavelength of the laser beam; spatial resolution is about 40 μm, which is determined by the diameter of the TM of 5 mm divided by 120 pixels; resolution of the displacement amplitude is 1.47 nm, which is determined by the maximum measurable displacement amplitude of 120 nm (peak to peak) divided by 82 digitized intensity levels of the frame memory; the noise floor is less than 10 nm in the frequency range from 0.5 to 10 kHz; the system is verified to work correctly up to 9 kHz by examining vibration modes of the circular rubber membrane.

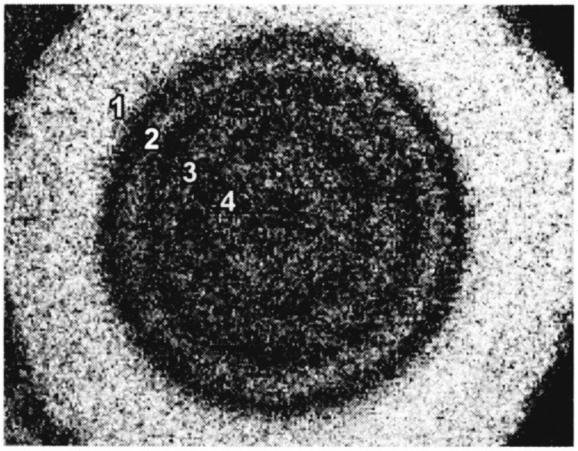
The measurements mentioned above are repeated five times and averaged to obtain vibration modes. In order to obtain the standard deviation (s.d.) in the displacement amplitude, this procedure is repeated ten times in one specimen. When sound-pressure levels are between 70 and 85 dB SPL and frequencies are between 2.0 and 3.5 kHz, the s.d. is ± 1.5 nm. However, when frequency responses are obtained, measurements are performed only once at intervals of 0.1 kHz in the frequency range from 1.0 to 3.0 kHz, as such measurements are time-consuming. In this case, the s.d. is ± 5.0 nm after measurements have been done ten times.

Twelve albino guinea pigs weighing between 250 and 400 g were employed. They were anesthetized by intraperitoneal injection of 35-mg/kg pentobarbital sodium and were decapitated under sedation. Temporal bones, which included air spaces of the middle ear and recesses that housed ossicles and muscles, were then removed. Cochleae were retained to be intact. Bony ear canals of the temporal bones were drilled to expose TMs. Finally, the temporal bones, except for the TMs, were embedded in plaster. Although it was not confirmed that the static pressure within the cavity was equalized to the ambient pressure, equalization would have been realized, owing to the existence of the small osseous opening of the Eustachian tube, when the temporal bones were removed. Measurements were performed immediately after the removal of the temporal bones. All the measurements were finished within 3 h after the removal of the temporal bones. During the measurements, saline was sprayed onto the TM to prevent it from drying. In order to increase the intensity of the reflected beam, white powder was spread over the surface of the TM. Care and use of animals reported in this study were approved by the Institutional Animal Care and Use Committee of Tohoku University.

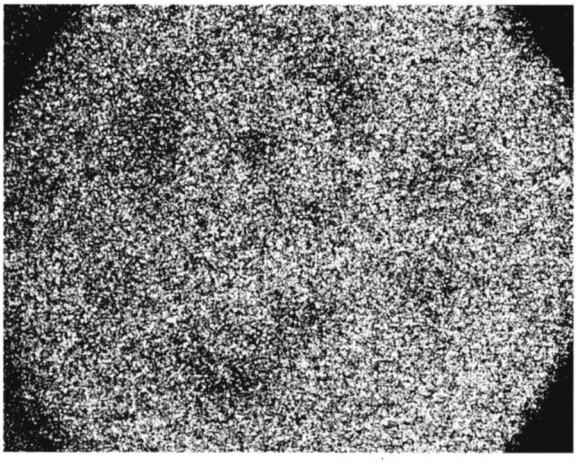
IV. RESULTS

A. Comparison between conventional SPI and SPI combined with the SPM

As shown in Fig. 2, the intensity at each pixel of the image obtained from conventional SPI changes with an increase in the displacement amplitude of the object vibrations. Figures 5(a) and (b) show the interference speckle patterns of the vibration of a circular rubber membrane 10 mm in diameter with the large and small amplitude, respectively, when conventional SPI is applied. Each bright numbered fringe shown in Fig. 5(a) corresponds to the number of the intensity peaks depicted in Fig. 2. However, as shown in Fig. 5(b), when the displacement amplitude is small, the vibration mode is unclear due to the small intensity variation shown in Fig. 3.



(a)



(b)

FIG. 5. Interference speckle patterns of the vibration of a circular rubber membrane 10 mm in diameter when conventional SPI is applied. (a) Large amplitude. The input sound-pressure level P is 75 dB SPL and its frequency f is 1.9 kHz. Each bright numbered fringe corresponds to the number of the intensity peaks depicted in Fig. 2. (b) Small amplitude. $P=90$ dB SPL and $f=6.25$ kHz. When the displacement amplitude is small, the vibration mode is not clear, due to the small intensity variation shown in Fig. 3.

Figure 6 shows the amplitude and phase of the circular rubber membrane obtained from Eqs. (14) and (15) with the same small amplitude and frequency as those shown in Fig. 5(b) when SPI combined with the SPM is applied. In contrast with the interference speckle pattern obtained by conventional SPI, which is shown in Fig. 5(b), clear information on the amplitude and phase is seen, in spite of the small amplitude. Substituting the data shown in Fig. 6 into Eq. (16), perspective plots of the displacement distribution of the circular rubber membrane are obtainable and displayed in Fig. 7 when $n=1$ and 3.

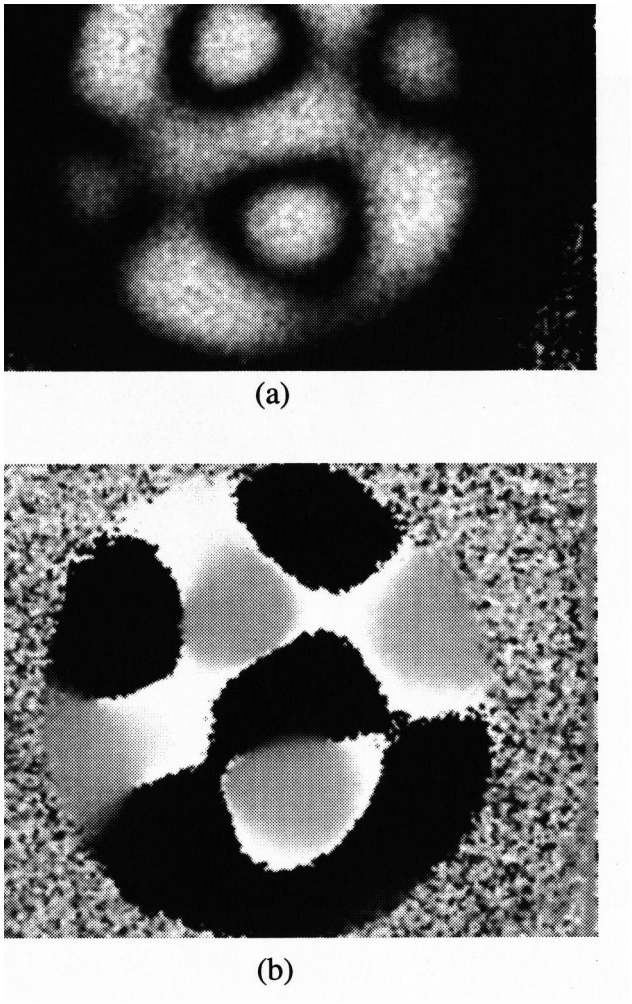


FIG. 6. Amplitude and phase of the circular rubber membrane obtained using SPI combined with the SPM. Stimulus condition is the same as that shown in Fig. 5(b). (a) Amplitude. The amplitudes in the white and black areas correspond to 60 and 0 nm, respectively, and the amplitude in between is expressed by gradation. (b) Phase. The phases in the white and black areas correspond to π and $-\pi$, respectively, and the phase in between is expressed by gradation.

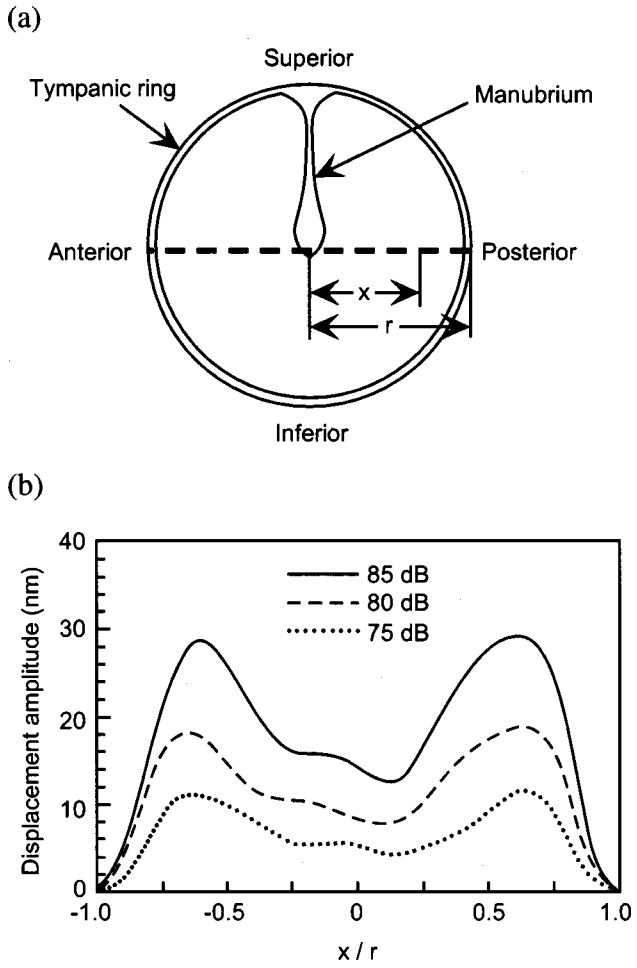


FIG. 8. Vibration amplitude of the TM. (a) Measurement position. Measurement is carried out at the dashed line across the umbo. (b) Distribution of the displacement amplitude when the input sound-pressure levels are 75, 80, and 85 dB SPL and its frequency is 1.0 kHz. The amplitude increases, but the mode is not altered with an increase in the sound-pressure level.

B. Displacement distribution of the TM

Although measurement was conducted in 12 guinea pigs, data were not obtained in the same condition, i.e., at the same stimulus level and frequency. Therefore, the results shown below come from one specimen.

Figure 8(b) shows the distribution of the displacement amplitude of the TM vibrations at the dashed line across the umbo shown in Fig. 8(a) when the input sound-pressure levels are 75, 80, and 85 dB SPL and its frequency is 1.0 kHz. The amplitude increases, but the mode does not alter with an increase in the sound-pressure level. The relationship between the displacement amplitude of the TM vibrations at the umbo and the input sound-pressure level, which is obtained from Fig. 8, is displayed in Fig. 9. The figure reveals that the displacement amplitude increases linearly with an increase in the sound-pressure level.

Figures 10 and 11 show the perspective plots and con-

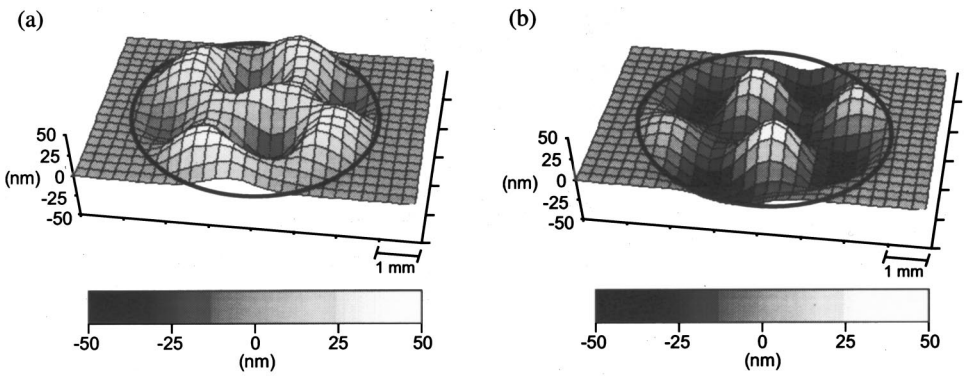


FIG. 7. Perspective plots of the displacement distribution of the circular rubber membrane using the amplitude and phase images shown in Fig. 6. (a) $n=1$. (b) $n=3$.

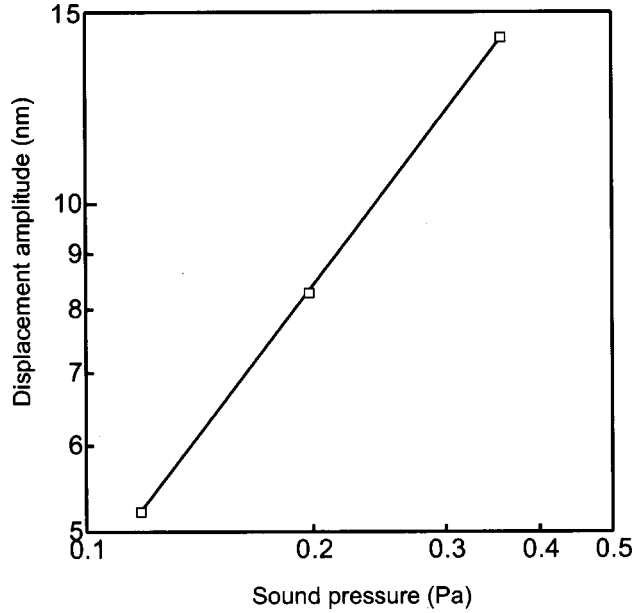


FIG. 9. Relationship between the displacement amplitude of the TM vibrations at the umbo and the input sound-pressure level, which is obtained from Fig. 8. Displacement amplitude increases linearly with an increase in the sound-pressure level.

tour maps of the displacement distribution of the left TM vibrations, respectively, when the displacement amplitude has its maximum value. At the frequency of 1.0 kHz, the whole TM vibrates in phase. The maximum displacement amplitude in the posterior portion is 28.5 nm, which is nearly the same as that in the anterior portion, i.e., 28.0 nm. At the frequency of 2.5 kHz, the TM has two local maxima in the posterior and inferior portions. The phase difference between the two portions is about 180 deg. At the frequency of 3.0 kHz, four peaks around the manubrium are seen. The TM in one area has an opposite-phase amplitude to that in the next area. The number of peaks increases and the vibration mode becomes more complicated with an increase in frequency.

C. Effect of modifying the bulla impedance

Figure 12 shows the method of modifying the bulla impedance. After measuring the TM vibrations under the condition of the closed bulla, the temporal bone is taken out of the plaster. A hole 6.0 mm in diameter is drilled in the wall of the bulla, and one end of a rubber ventilation tube 1.0 m in length is attached to the hole by glue. The other end is open into the free field. The temporal bone is then fixed again by plaster.

Figures 13 and 14 show perspective plots of the displacement distribution of the TM vibrations and the distribution of the displacement amplitude at the dashed line across the umbo, before and after modifying the bulla impedance, when $f=1.0$ kHz and $P=85$ dB SPL. The vibration mode after modifying the bulla impedance is similar to that in the case of the closed bulla. However, the maximum value of the displacement amplitude increases 9.9% after modifying the bulla impedance. As shown in Figs. 15 and 16, when $f=2.5$ kHz and $P=70$ dB SPL, before modifying the bulla impedance, the vibration pattern has two local peaks in the posterior and inferior portions and the phase difference between them is noticeable. By contrast, after modifying the bulla impedance, although the vibration pattern also has two

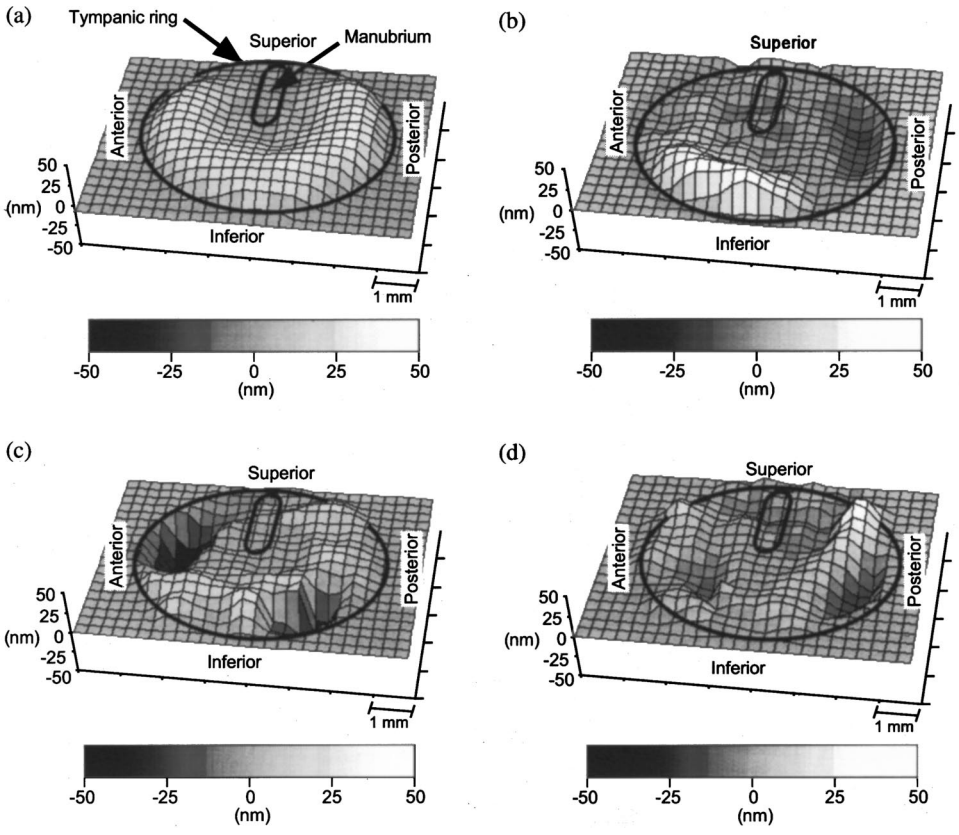


FIG. 10. Perspective plots of the displacement distribution of the left TM vibrations, when the displacement at each point reaches its maximum value. (a) Frequency $f=1.0$ kHz and sound-pressure level $P=85$ dB SPL. (b) $f=2.5$ kHz and $P=70$ dB SPL. (c) $f=3.0$ kHz and $P=75$ dB SPL. (d) $f=4.0$ kHz and $P=75$ dB SPL.

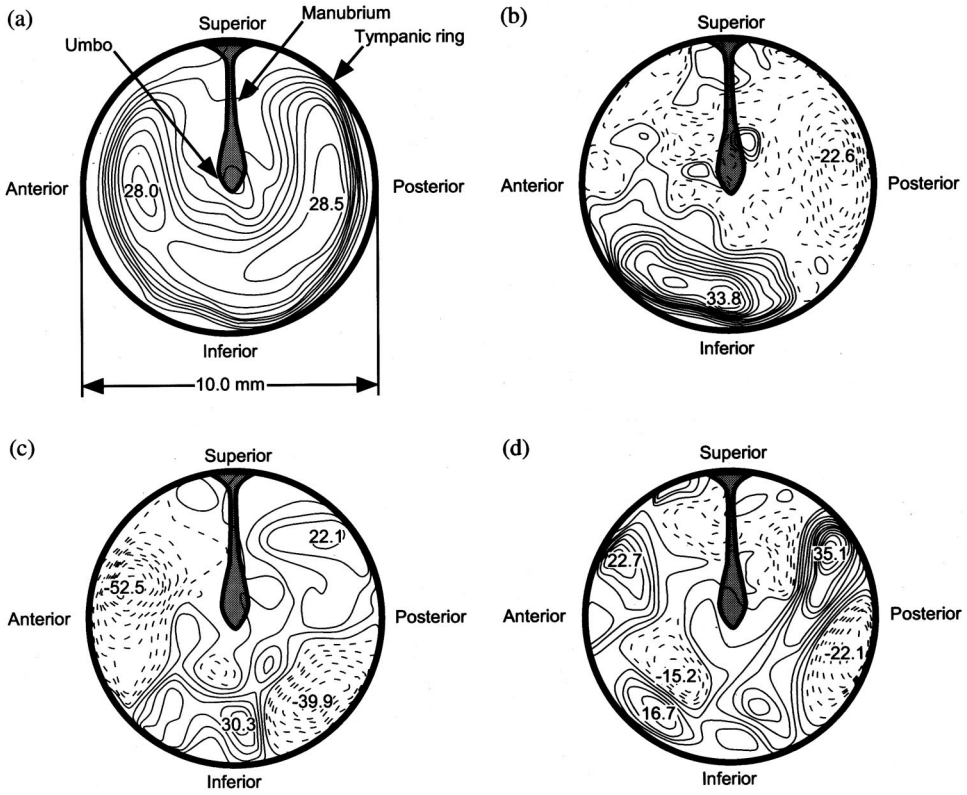


FIG. 11. Contour maps corresponding to Fig. 10. The contours represent lines of the constant vibration amplitude. Intervals are 3 nm in (a), (b), and (d), and 5 nm in (c). The solid and dotted contours represent the upward and downward displacements, respectively. Also, the positive and negative digits are the upward and downward maximum values of the displacements at the specific points, respectively.

local peaks in the posterior and inferior portions, there is no phase difference between them.

Figures 17(b) and (c) show the frequency responses of the displacement amplitude at the middle of the inferior portion and at the umbo, respectively, marked by the open square and the $\times$ in Fig. 17(a). In these measurements, the frequency responses are obtained at intervals of 0.1 kHz in the frequency range from 1.0 to 3.0 kHz and $f=0.5$, 0.6, 0.8, 0.9, 3.3, 3.5, and 4.0 kHz. At the middle of the inferior portion, after modifying the bulla impedance, the displacement amplitude increases in the frequency region below 2.1 kHz, but the change is slight in the frequency region above 2.1 kHz. In the umbo as well, the displacement amplitude increases in the low-frequency region after modifying the bulla impedance. The resonance frequency, i.e., the frequency where the umbo has a maximum displacement amplitude, marked by the open triangles in Fig. 17(c), decreases from 2.7 to 2.3 kHz after modifying the bulla impedance.

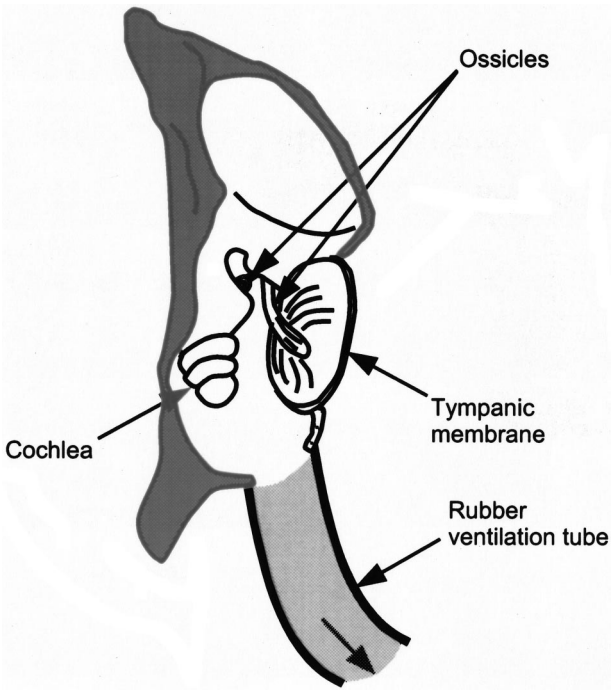


FIG. 12. Method of modifying the bulla impedance. A hole 6.0 mm in diameter is drilled in the wall of the bulla, and one end of a rubber ventilation tube 1.0 m in length is attached to the hole by glue. The other end is open into the free field.

V. DISCUSSION

A. Measurement system

Using conventional SPI, Khanna and Tonndorf (1972), and Tonndorf and Khanna (1972) beautifully showed us the vibration pattern of the TMs in cats and human cadavers. However, to obtain clear fringes a high sound pressure, which was larger than 100 dB SPL, had to be supplied as stimulation. Also, phase information of TM vibration modes was not obtained.

For the purpose of overcoming these two major weak points, a new method, i.e., time-averaged SPI combined with the SPM, was developed (Johns and Wykes, 1989). Using this method, Hogmonen and Lokberg (1977) measured the TM frequency response at the malleus in a human temporal bone, when $P=90$ dB SPL. However, as they used a photodiode with an active area of 100 mm^2 to measure the displacement amplitude of the TM, they could not obtain the vibration mode of the TM. In their related paper' (Lokberg *et al.*, 1980), they recorded the TM frequency responses at

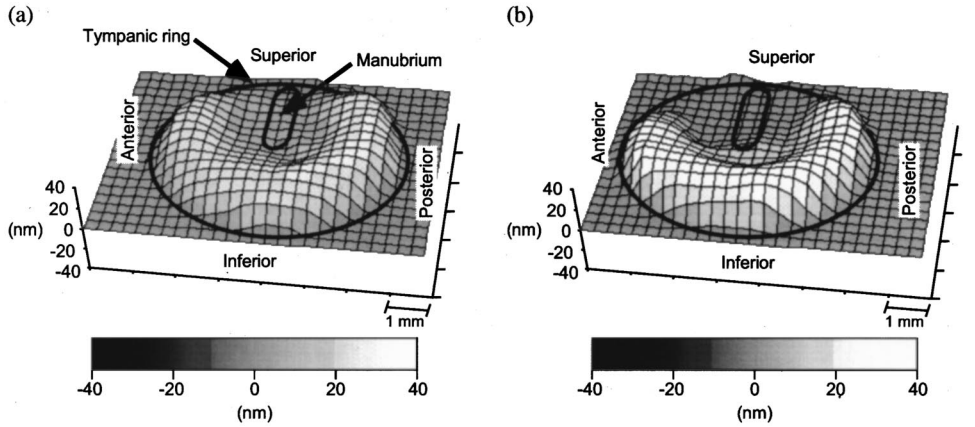


FIG. 13. Comparison of the displacement distribution of the TM vibrations in the case of the closed bulla with that after modifying the bulla impedance when $f=1.0$ kHz and $P=85$ dB SPL. (a) Closed bulla. (b) After modifying the bulla impedance. The difference in the vibration mode between before and after modifying the bulla impedance is small.

the malleus manubrium in humans *in vivo*, and sketched the vibratory behavior of the TM. However, it was only when $f=2.4$ kHz and $P=90$ dB SPL.

In their experiments, phase modulation of the reference beam was provided mechanically by a modulating mirror, which was mounted on a piezoelectric drive element. By contrast, in our system, owing to the recent developments in electronics, a phase modulation of the reference beam can be realized by a direct modulation of laser injection current. The electrical phase modulation of the reference beam would be superior to that achieved by the mechanical modulation of the mirror, especially in the high-frequency region because the inertia effect of the mirror must be taken into account when the mirror is mechanically vibrated. Also, with recent advances in digital image processing techniques and by use of the procedure of raising the resolution of the displacement amplitude, as shown in Figs. 10, 11, 13, 15, and 16, clear perspective plots and detailed contour maps of the displacement distribution of the TM vibrations are obtainable. Furthermore, realistic dynamic animation of the TM vibrations is possible, an example of which can be seen at our web site: http://www.wadalab.mech.tohoku.ac.jp/.

A commercially available scanning LDV is a sensitive, noncontacting device capable of measuring the displacement amplitude of TM vibrations in the nanometer range. It has been already used to explore the function of the middle ear in human cadaveric temporal bones and live human subjects (Konradsson *et al.*, 1987; Goode *et al.*, 1996; Ball *et al.*, 1997; Huber *et al.*, 2001a, b). A built-in camera and a computer-driven mirror system control the laser position on the object with a spatial resolution of 5 μm and no reflection material is needed on the object. However, as it is a scanning system, in order to obtain the vibration pattern of the object, many points on the object, e.g., 95 points on the TM (Huber *et al.*, 2001a), must be analyzed. As a result, it takes 50 to 90 s to acquire the measurement data depending on the number of scanning points selected. By contrast, in our system, although white powder must be spread over the surface of the TM in order to increase the intensity of the reflected beam, scanning is not necessary and the data in the whole area of the object are obtained simultaneously in 1 s. In spite of such a short time required for the acquisition of data, the resolution of our system is of nanometer order, the same as that of a scanning LDV system.

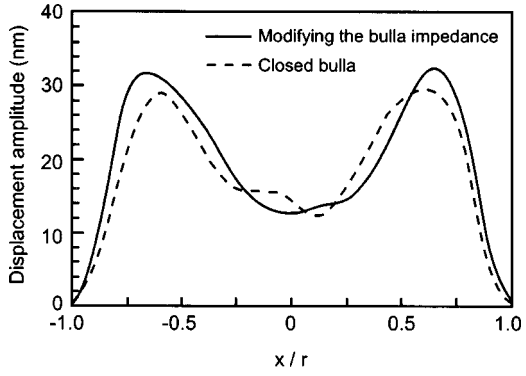


FIG. 14. Distribution of the displacement amplitude of the TM at the dashed line across the umbo, which is shown in Fig. 8(a), before and after modifying the bulla impedance. Stimulus condition is the same as that shown in Fig. 13. The maximum value of the displacement amplitude increases 9.9% after modifying the bulla impedance.

B. Dynamic behavior of the TM

First, it must be pointed out that it is not possible to discuss the difference in results among specimens, as the results shown so far were obtained from one guinea pig.

To date, the steady-state vibration patterns of the TM in the cat (Khanna and Tonndorf, 1972), the frog (Kimura, 1981), the human temporal bone (Tonndorf and Khanna, 1972) and in the human *in vivo* (Lokberg *et al.*, 1980), and the transient vibration patterns of the TM in the guinea pig (Dancer *et al.*, 1975) have been recorded. The whole TM vibrates in phase when $f<1.0$ kHz, and as shown in Fig. 8(b), the distribution of the displacement amplitude in the posterior region is nearly the same as that in the anterior region. On the contrary, in cats and humans, the maximum in the posterior region is larger than that in the anterior region. This inconsistency would come from the difference in the configuration. As shown in Fig. 8(a), the manubrium of the guinea pig is located so as to divide the TM equally into the posterior and anterior regions (Dancer *et al.*, 1975; Wada *et al.*, 1998). By contrast, the manubrium in the cat and human does not divide the TM evenly, and the area of the posterior region is larger than that of the anterior region (Kirikae, 1960; Decraemer and Dirckx, 1991; Decraemer, Dirckx, and Funnell, 1991).

The middle ear of the cat was found to manifest the linear behavior even when the high sound stimulation be-

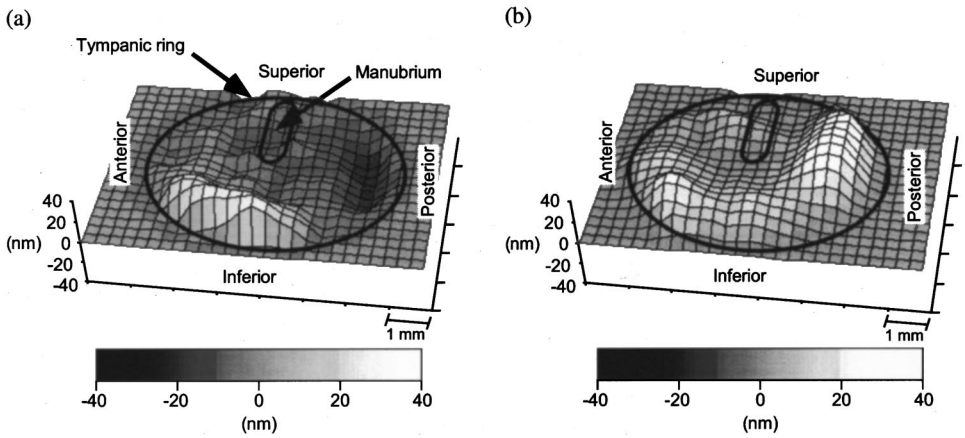


FIG. 15. Comparison of the displacement distribution of the TM vibrations in the case of the closed bulla with that after modifying the bulla impedance when $f=2.5$ kHz and $P=70$ dB SPL. (a) Closed bulla. (b) After modifying the bulla impedance.

tween 90 and 113 dB SPL was supplied (Khanna and Tonndorf, 1972). Therefore, it is reasonable that the displacement amplitude of the TM of the guinea pig increases linearly without alternating its mode when the stimulus sound pressure increases from 70 to 85 dB SPL.

C. Effect of the bulla on the dynamic behavior of the TM

From our experimental results, it can be said that by modifying the bulla impedance the displacement amplitude of the TM vibration along the line across the umbo increases less than 10% without an alteration of its mode when the stimulus frequency is 1.0 kHz. The results also reveal that the vibration mode of the TM is changed by modifying the bulla impedance without increasing the displacement amplitude, when the stimulus frequency is around 2.5 kHz.

In our previous measurement (Wada *et al.*, 1991), frequency characteristics of the middle ear changed drastically below 3.0 kHz, and the middle-ear resonance frequency decreased from 2.5 to 1.1 kHz after opening of the bulla. Zwislocki (1963) also showed that the magnitude of the impedance at the TM was altered from the frequency between 2.0 and 3.0 kHz to the frequency around 0.8 kHz after opening of the bulla. Although the species was different from ours, the measurement data of Ruggero *et al.* (1990) revealed that the peak velocity of the manubrium mallei of the chinchillas increased nearly ten times and that the resonance frequency decreased from 2.0 to 0.2 kHz after opening of the bulla. However, our experimental data, which are shown in Figs. 17(b) and (c), indicate that the displacement amplitude of the TM increases only between two and three times and that the resonance frequency decreases only from 2.7 to 2.3 kHz after modifying the bulla impedance.

In our previous experiment (Wada *et al.*, 1991), the bulla was fully opened, and the stiffness of the air within the bulla was reduced nearly to zero. By contrast, as shown in Fig. 12, in this experiment, by gluing a long rubber ventilation tube 6.0 mm in diameter and 1.0 m in length to the wall of the bulla, in other words, by modifying the bulla impedance, the condition of the open bulla is realized. As a result, the stiffness of the air within the bulla and tube would affect the dynamic behavior of the TM.

The simplified impedance of the middle ear Z and its resonance frequency f_{RES} are expressed in the following form:

$$Z=C+j(2\pi fM)-j(K/2\pi f), \tag{17}$$

$$f_{\mathrm{RES}}=(1/2\pi)(K/M)^{1/2}, \tag{18}$$

where C, M, and K are the damping, mass, and stiffness of the middle ear, respectively, and $j=(-1)^{1/2}$. Equation (17) shows us that the change in the stiffness component of the impedance Z is large in the low-frequency region. Therefore, in our experiment, the reason why the increase in the displacement amplitude of the TM vibration is suppressed in the low-frequency region would be due to the relatively large stiffness of the bulla and tube after modifying the bulla impedance. From Eq. (18) as well, it is understandable that the

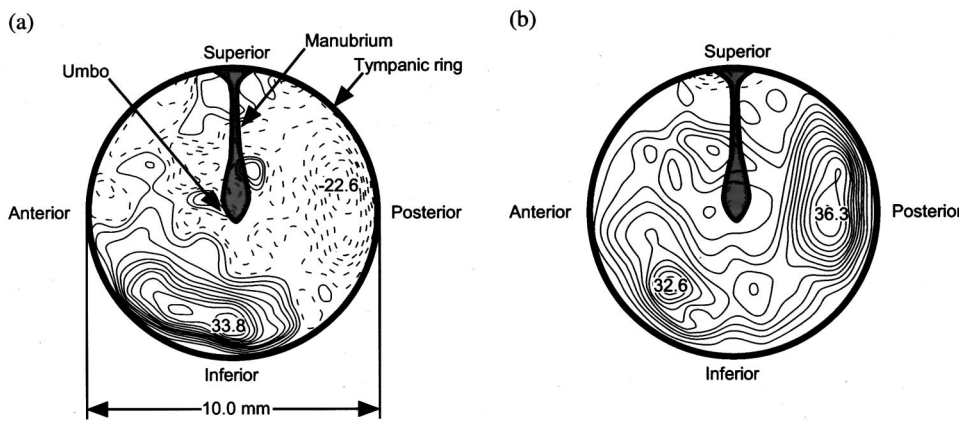


FIG. 16. Contour maps corresponding to Fig. 15, when $f=2.5$ kHz and $P=70$ dB SPL. The contours represent lines of the constant vibration amplitude, and intervals are 3 nm. The solid and dotted contours represent the upward and downward displacements, respectively. Also, the positive and negative digits mean the upward and downward maximum values of the displacements at the specific points, respectively.

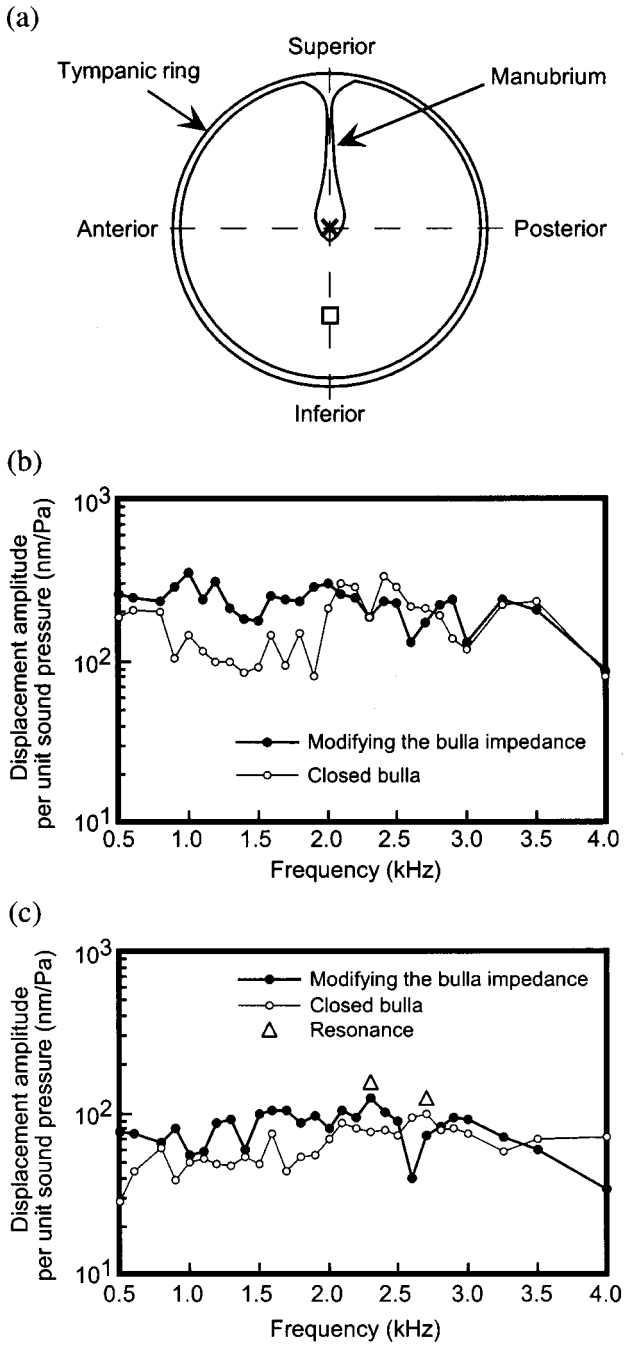


FIG. 17. Vibration amplitude of the TM. (a) Measurement points. (b) Frequency responses of the displacement amplitude at the middle of the inferior portion, which is marked by the open square in (a). (c) Frequency responses of the displacement amplitude at the umbo, which is marked by the $\times$ in (a). The resonance frequency, which is marked by the open triangles, decreases from 2.7 kHz to 2.3 kHz after modifying the bulla impedance.

drastic decrease in the resonance frequency after modifying the bulla impedance is not seen in our experiment, due to the same reason mentioned above.

In addition to a shift of the resonance frequency, a sharp notch was observed at 2.6 kHz after modifying the bulla impedance. Zwislocki (1963) showed that the impedance at the TM with the bulla open had a maximum value at approximately 2.6 kHz, which was not observed in the intact ear, and that this phenomenon was caused by the change in the impedance of the bulla. Therefore, although it is not obvious, the sharp notch at 2.6 kHz may be caused by the antiresonance, i.e., a local rapid increase in the impedance, of the middle ear which was induced by modifying the bulla impedance.

VI. CONCLUSIONS

Using time-averaged electronic SPI, the SPM technique by a direct modulation of laser injection current and the digital image processing technique, an attempt was made to develop a system for measuring a small vibration sensitively. The vibratory responses of the TM of guinea pig temporal bones were then measured by this system. The conclusions are as follows.

(1) The system herein presented enables us to measure the nanometer-order displacement amplitude of the TM vibration produced by stimuli of 70 dB SPL, and to obtain phase information on the TM vibration. Furthermore, realistic animation of the TM vibrations is made possible, which facilitates understanding of the dynamic behavior of the TM.
(2) At the frequency of 1.0 kHz, the whole TM vibrates in phase. At the frequency of 2.5 kHz, the TM has two opposite phase maxima in the posterior and inferior portions. The number of the peaks then increases and the vibration mode becomes more complicated with an increase in frequency.
(3) When the bulla impedance is modified by attachment of a rubber ventilation tube to a hole in the wall of the bulla, the displacement amplitude of the TM vibration increases without an alteration of its mode when the stimulus frequency is 1.0 kHz. However, when $f = 2.5$ kHz, the mode of the TM changes without an increase in displacement amplitude after modifying the bulla impedance.

ANSI (**1987**). ANSI S.3.39-1987, "Specifications for instruments to measure aural acoustic impedance and admittance" (American National Standards Institute, New York).
Ball, G. R., Huber, A., and Goode, R. L. (**1997**). "Scanning laser Doppler vibrometry of the middle ear ossicles," Ear Nose Throat J. **76**, 213–216, 218, 220, 222.
Buunen, T. J. F., and Vlaming, M. S. M. (**1981**). "Laser-doppler velocity meter applied to tympanic membrane vibrations in cat," J. Acoust. Soc. Am. **69**, 744–750.
Dancer, A. L., Franke, R. B., Smigielski, P., Albe, F., and Fagot, H. (**1975**). "Holographic interferometry applied to the investigation of tympanic-membrane displacements in guinea pig ears subjected to acoustic impulses," J. Acoust. Soc. Am. **58**, 223–228.
Decraemer, W. F., and Dirckx, J. J. (**1991**). "Moire-shift interferometer measurements of the shape of human and cat tympanic membrane," Proc. SPIE **1429**, 26–33.
Decraemer, W. F., Dirckx, J. J., and Funnell, W. R. J. (**1991**). "Shape and derived geometrical parameters of the adult, human tympanic membrane measured with a phase-shift moire interferometer," Hear. Res. **51**, 107–122.
Decraemer, W. F., Khanna, S. M., and Funnell, W. R. J. (**1989**). "Interferometric measurement of the amplitude and phase of tympanic membrane vibrations in cat," Hear. Res. **38**, 1–18.
Decraemer, W. F., Khanna, S. M., and Funnell, W. R. J. (**1991**). "Malleus vibration mode changes with frequency," Hear. Res. **54**, 305–318.
Ellingsrud, S., and Rosvold, G. O. (**1992**). "Analysis of a data-based TV-holography system used to measure small vibration amplitudes," J. Opt. Soc. Am. **9**, 237–251.

Goode, R. L., Ball, G., and Nishihara, S. (**1993**). ''Measurement of umbo vibration in human subjects: Method and possible clinical applications,'' Am. J. Otol. **14**, 247–251.
Goode, R. L., Ball, G., Nishihara, S., and Nakamura, K. (**1996**). ''Laser Doppler vibrometer (LDV): A new clinical tool for the otologist,'' Am. J. Otol. **17**, 813–822.
Hogmonen, K., and Lokberg, O. J. (**1977**). ''Detection and measurement of small vibrations using electronic speckle pattern interferometry,'' Appl. Opt. **16**, 1869–1875.
Huber, A., Schwab, C., Linder, T., Stoeckli, S., Ferrazzini, M., Dillier, N., and Fisch, U. (**2001a**). ''Evaluation of eardrum laser Doppler interferometry as a diagnostic tool,'' Laryngoscope **111**, 501–507.
Huber, A., Linder, T., Ferrazzini, M., Schmid, S., Dillier, N., Stoeckli, S., and Fisch, U. (**2001b**). ''Intraoperative assessment of stapes movement,'' Ann. Otol. Rhinol. Laryngol. **110**, 31–35.
Johns, R., and Wykes, C. (**1989**). *Holographic and Speckle Interferometry* (Cambridge University Press, Cambridge).
Khanna, S. M., and Tonndorf, J. (**1972**). ''Tympanic membrane vibrations in cats studied by time-averaged holography,'' J. Acoust. Soc. Am. **51**, 1904–1920.
Kimura, Y. (**1981**). ''Experimental study of the vibration analysis of tympanic membrane by holographic interferometry,'' J. Otolaryngol. Jpn. **84**, 880–889.
Kirikae, I. (**1960**). *The Structure and Function of Middle Ear* (Tokyo University Press, Tokyo), pp. 37–38.
Konradsson, K. S., Ivarsson, A., and Bank, G. (**1987**). ''Computerized laser Doppler interferometric scanning of the vibrating tympanic membrane,'' Scand. Audiol. **16**, 159–166.
Lilly, D. J. (**1984**). ''Multiple frequency, multiple component tympanometry: New approaches to an old diagnostic problem,'' Ear Hear. **5**, 300–308.
Lokberg, O. J., and Hogmonen, K. (**1976**). ''Vibration phase mapping using electronic speckle pattern interferometry,'' Appl. Opt. **15**, 2701–2704.
Lokberg, O. J., Hogmonen, K., and Gundersen, T. (**1980**). ''Vibration measurement of the human tympanic membrane: In vivo,'' Acta Oto-Laryngol. **89**, 37–42.
Nishihara, S., Aritomo, H., and Goode, R. L. (**1993**). ''Effect of changes in mass on middle ear function,'' Otolaryngol.-Head Neck Surg. **109**, 899–910.
Petzing, J. N., and Tyrer, J. R. (**1998**). ''Recent developments and applications in electronic speckle pattern interferometry,'' J. Strain Anal. **33**, 153–169.
Ruggero, M. A., Rich, N. C., Robles, L., and Shivapuja, B. G. (**1990**). ''Middle-ear response in the chinchilla and its relationship to mechanics at the base of the cochlea,'' J. Acoust. Soc. Am. **87**, 1612–1629.
Saldner, H. O. (**1996**). ''Phase-stepped television holographic technique for measuring phase and amplitude maps of small vibrations,'' Appl. Opt. **35**, 3791–3798.
Tonndorf, J., and Khanna, S. M. (**1972**). ''Tympanic-membrane vibrations in human cadaver ears studied by time-averaged holography,'' J. Acoust. Soc. Am. **52**, 1221–1233.
Vlaming, M. S. M. G., and Feenstra, L. (**1986a**). ''Studies on the mechanics of the normal human middle ear,'' Clin. Otolaryngol. **11**, 353–363.
Vlaming, M. S. M. G., and Feenstra, L. (**1986b**). ''Studies on the mechanics of the reconstructed human middle ear,'' Clin. Otolaryngol. **11**, 411–422.
von Bekesy, G. (**1949**). ''The structure of the middle ear and the hearing of one's own voice by bone conduction,'' J. Acoust. Soc. Am. **21**, 217–232.
Wada, H., and Kobayashi, T. (**1990**). ''Dynamical behavior of middle ear: Theoretical study corresponding to measurement results obtained by a newly developed measuring apparatus,'' J. Acoust. Soc. Am. **87**, 237–245.
Wada, H., Kobayashi, T., Ohyama, K., and Nishimura, S. (**1991**). ''Frequency characteristics of guinea pig and human middle ear up to 4 kHz,'' J. Acoust. Soc. Jpn. **47**, 235–242.
Wada, H., Momose, K., Sugawara, M., Hozawa, K., and Takasaka, T. (**1998**). ''Measurement of guinea pig middle ear using computer-aided three-dimensional reconstruction system,'' Otol. Jpn. **8**, 511–515.
Zwislocki, J. (**1963**). ''Analysis of the middle-ear function. II. Guinea-pig ear,'' J. Acoust. Soc. Am. **35**, 1034–1040.

Re-examination of the role of the human acoustic stapedius reflex

Dennis P. Phillips[a)]
Hearing Research Laboratory, Department of Psychology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

Andrew Stuart and Michael Carpenter
Department of Communication Sciences and Disorders, East Carolina University, Greenville, North Carolina 27858-4353

(Received 18 October 2001; revised 11 February 2002; accepted 12 February 2002)

The ''rollover'' seen in the word recognition performance scores of patients with Bell's palsy (facial nerve paralysis) has historically been taken as an indicator of the role of the stapedius reflex in the protection from upward spread of masking. Bell's palsy, however, may be a polyneuropathy, so it is not clear that the poor word recognition performance at high levels is necessarily attributable specifically to impaired facial nerve function. The present article reports two new experiments that probe whether an isolated impairment of the stapedius reflex can produce rollover in word recognition performance-intensity functions. In experiment 1, performance-intensity functions for monosyllabic speech materials were obtained from ten normal listeners under two listening conditions: normal and low-frequency augmented to offset the effects of the stapedius reflex on the transmission of low-frequency vibrations to the cochlea. There was no effect of the spectral augmentation on word recognition for stimulus levels up to 107 dB SPL. In experiment 2, six patients who had undergone stapedectomy were tested for rollover using performance-intensity functions. None of the patients showed rollover in their performance-intensity functions, even at stimulus levels in excess of 100 dB HL. These data suggest that if the stapedius reflex has a role in protection from upward spread of masking, then this role is inconsequential for word recognition in quiet. [DOI: 10.1121/1.1467670]

PACS numbers: 43.64.Ha, 43.66.Lj, 43.71.Es, 43.71.Ky [BLM]

I. INTRODUCTION

The contraction of the stapedius muscle, whether brought about by intense external sounds (e.g., Borg, 1968) or by internally generated vocalization (Borg and Zakrisson, 1975), results in a reduction in the transmission of low-frequency vibrations to the cochlea (Nuttall, 1974; Wilson and Margolis, 1991). One main hypothesis of the function of the acoustic stapedius reflex is based on this high-pass filtering action, *viz.*, that the reflex protects the perception of high-frequency signals from masking by intense low-frequency sounds or sound components (Stevens and Davis, 1938; Liden *et al.*, 1963; Borg and Zakrisson, 1974, 1975; Mahoney *et al.*, 1979; Borg and Counter, 1989).

One independent line of evidence adduced in support of this hypothesis has been the word recognition performance of patients with inactive stapedius reflexes secondary to facial nerve paralysis (patients with ''Bell's palsy:'' Borg and Zakrisson, 1973, 1974; McCandless and Schumacher, 1979). These listeners display word recognition which is worse for very high amplitude signals (roughly, those above stapedius reflex threshold) than their own performance with lower-amplitude sounds (''rollover,'' after Jerger and Jerger, 1971). Patients with facial nerve paralysis may also show disturbances in loudness perception for sounds whose amplitudes exceed reflex threshold (McCandless and Goering, 1974; McCandless and Schumacher, 1979; Phillips and Carr, 1998). The argument has thus been that in the absence of the stapedial contractions, high-amplitude sounds result in greater-than-normal low-frequency energy being transmitted to the cochlea, and this results in both exaggerated loudness and a masking of high-frequency stimulus elements.

The difficulty with this line of evidence is that Bell's palsy may be a polyneuropathy of cranial nerve systems, including the auditory one (Adour, 1975; Citron and Adour, 1978; Lundgren *et al.*, 1977; Phillips and Carr, 1998). Other authors have found auditory brainstem response abnormalities in Bell's palsy patients, but in the absence of cochlear nerve pathology (Rosenthal *et al.*, 1983). Accordingly, the rollover in the word recognition scores of patients with Bell's palsy may be due to neurological factors other than a paralyzed stapedius muscle. In this regard, there is evidence that patients with or without intact stapedius reflexes, but with cochlear nerve or brainstem disease, also often display word recognition rollover (Jerger and Jerger, 1971; Dirks *et al.*, 1977; Hannley and Jerger, 1981; Meyer and Mishler, 1985). Other authors, specifically studying the auditory responses of patients with Bell's palsy, have found little or no evidence of cochlear nerve or auditory brainstem response abnormalities (Hendrix and Melnick, 1983; Maurizi *et al.*, 1987). One group studying Bell's palsy has found evidence of decreased reflex strength at high stimulus levels, but in the absence of

a)Author to whom correspondence should be addressed. Address for all correspondence: Dr. D. P. Phillips, Hearing Research Laboratory, Department of Psychology, Dalhousie University, Halifax, NS B3H 4J1, Canada; electronic mail: ears@is.dal.ca

reflex threshold elevations (Ardic *et al.*, 1997). Taken together, these data suggest that it may be problematic to use data from patients with Bell's palsy to isolate the role of the stapedius reflex in preventing rollover in word recognition unless there is also careful confirmation of the absence of concurrent auditory brainstem abnormalities. The basic question which remains, then, is whether isolated impairment of the stapedius reflex is able to produce rollover in word recognition performance-intensity functions.

One way of studying this issue has involved simulation of a paralyzed stapedius in normal listeners. Wormald *et al.* (1995) obtained word recognition performance-intensity functions in patients with Bell's palsy, and word recognition performance-intensity functions with low-frequency enhancement in normal listeners. Rollover was measured at 100 dB HL. The normal listeners showed rollover in response to the enhanced speech, but this rollover was less than that seen in patients with Bell's palsy studied by the same group. This led the authors to conclude that a paralyzed stapedius alone could not account for the rollover in word recognition scores of patients with Bell's palsy. By extension, this conclusion weakens the case for a role of the stapedius in protection from upward spread of masking.

Dorman *et al.* (1987; see also Dorman *et al.*, 1986) took a different approach. They studied the performance-intensity functions of normal listeners for 50-ms, synthetic vowels, i.e., speech samples shorter than the stapedius reflex latency. They found up to 50% rollover in the vowel identification scores for stimulus levels above about 90 dB SPL. These data suggest that the strictly *mechanical* effect of an absent stapedius reflex is sufficient to result in impoverished auditory encoding of vowels.

The present report describes two experiments that attempt to provide independent lines of evidence on the role of the stapedius muscle in the recognition of words in quiet. In experiment 1, the performance-intensity functions for speech materials of normal listeners were studied under stimulating conditions designed to mimic the absence of a stapedius reflex (i.e., an absent stapedius muscle). In experiment 2, stapedectomy patients (i.e., listeners with surgically confirmed severing of the stapedius attachment) were studied for their word recognition performance in quiet.

II. EXPERIMENT 1

The effects of stapedius contractions on middle-ear sound transmission have been described in detail (see Nuttall, 1974; Wilson and Margolis, 1991). We reasoned that it should be possible to reveal a role of the stapedius by offsetting the stapedius effects of the reflex through enhancement of the spectral content of the stimuli in reverse fashion. Experiment 1 is a study in which the performance-intensity functions of normal listeners for spoken words were examined for normal speech, and for low-frequency enhanced speech. In what follows, we have paid a great deal of attention both to simulating stapedial function accurately, and to careful scrutiny of the stimulus materials for unintended acoustical distortions.

A. Methods

1. Participants

Ten young adults ($M = 26.8$ years, s.d.$= 2.0$; nine females) served as participants. All participants presented with normal-hearing sensitivity, defined as having pure-tone thresholds at octave frequencies from 250 to 800 Hz, and speech recognition thresholds (SRTs), of $\leqslant$15 dB HL (ANSI, 1996). Participants also presented with normal middle-ear function by immittance audiometry (American Speech-Language-Hearing Association, 1997).

2. Stimuli and apparatus

The test stimuli consisted of ten isophonemic consonant–vowel–consonant (CVC) word lists developed by Olsen *et al.* (1997). Each list contained ten different words with the same 30 phonemes appearing in each list. Commercially available compact disc recordings of these lists were used (*Speech Audiometry Materials*, Brigham Young University, 1991). Ten of the stimuli were measured for their durations, and all were in the range from 650 to 800 ms.

A double-wall sound-treated audiometric suite (Industrial Acoustics Corporation), meeting specifications for permissible ambient noise (ANSI, 1999), served as the test environment. The recorded stimuli were presented to each participant through an insert earphone (Etymotic Research model ER-3A). The stimuli were routed from a compact disc player (Sony model CDP-CE 415) through a digital equalizer (Yamaha model DEQ5) to a clinical audiometer (Grason Stadler GSI 61 model 1761-9780XXE) and then to the earphone. All equipment was calibrated to manufacturer's and/or appropriate published standards prior to testing. During the normal listening conditions the speech signal was routed through the digital equalizer in a ''flat'' mode (i.e., unequalized). In the simulated stapedius muscle paralysis conditions, the speech signal was ''low-frequency enhanced'' to mimic the loss of the attenuating characteristics of the ipsilateral stapedius muscle contraction during the stapedius reflex.

The approximation of ipsilateral attenuation offered by the stapedius muscle to low-frequency acoustic energy during paralysis was based on data from Borg (1968), Borg and Zakrisson (1974), Rabinowitz (1977), and Zakrisson *et al.* (1974). To summarize, the following rationale was adopted: The stapedius reflex acts as a high-pass filter on the transmission of acoustic stimuli through the middle ear. The attenuation of low frequency occurs mostly below 2000 Hz, with the maximum effect around 600 Hz. The attenuation of low-frequency sound through the middle ear increases in direct proportion with increases in the level of the activator signal. Maximum attenuation (i.e., approximately 15 to 20 dB) can be expected to occur where ipsilateral reflex activity approaches its maximum (i.e., approximately 20 to 25 dB above the stapedius reflex threshold [ART]). As such, the maximum attenuation occurring at 600 Hz can be expected to increase roughly linearly with every dB increase in the activator sound-pressure level (SPL) to a maximum attenuation of approximately 20 dB at 25 dB above the ART. The slope of the roll-off of the high-pass filter increases from the

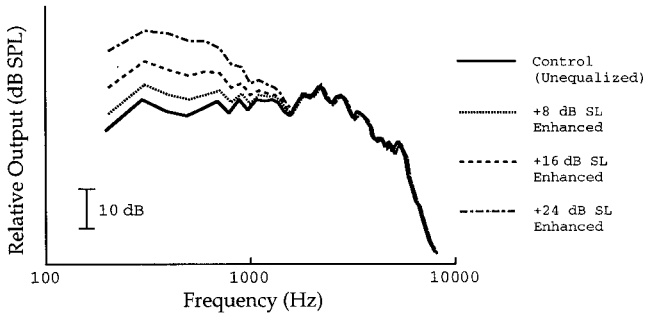


FIG. 1. Relative output (dB SPL), as a function of frequency (Hz), of the insert earphone during the unequalized control and low-frequency-enhanced listening conditions as measured in a 2-cm^3 HA-2 coupler.

knee point of 2000 Hz to the maximum point of attenuation with increases in the level of the activator signal above the ART (see Rabinowitz, as cited in Wilson and Margolis, 1991).

We attempted to simulate the effects of stapedius muscle paralysis for three sensation levels (SLs) above the ART of an otherwise normally functioning ear. The loss of stapedius muscle activity results in an increase in low-frequency acoustic energy reaching the cochlea. The digital equalizer was manipulated to give three levels of low-frequency enhancement above the ART (i.e., 8, 16, and 24 dB SL *re*: ART). Direct measurements of this low-frequency ipsilateral enhancement are illustrated in Fig. 1. A Fonix 6500 hearing aid test system (Frye Electronics) was employed to measure 2-cm^3-coupler SPL output of the insert earphone during the low-frequency enhanced listening conditions. A pure-tone sweep (200 Hz through 8000 Hz) served as stimulus. The stimulus was generated by the Fonix 6500 hearing aid test system and routed through the digital equalizer to the audiometer and delivered to the insert earphone. The insert earphone was coupled with 15 mm of #13 thick-walled tubing and an ear-level hearing aid adapter to a standard 2-cm^3 HA-2 coupler. Sampling was obtained between 200 and 8000 Hz in 100-Hz steps. It was assured that the 2-cm^3 coupler output in that range exceeded the noise floor of the hearing aid test system. Digital hard copies of the data were obtained. As evident in Fig. 1, the enhancement of low frequencies occurs below 2000 with the maximum effect between 300 and 600 Hz; enhancement of low frequencies increases in direct proportion with increases in SL above the ART; maximum enhancement (i.e., approximately 20 dB) occurs where ipsilateral reflex activity approaches its maximum (i.e., 24 dB above the ART); and the slope of the low-frequency enhancement increases from the knee point of 2000 Hz with increases in low-frequency enhancement for increasing levels in sensation above the ART.

In pilot testing, we became acutely aware of the problem of speech stimulus distortion in the playback chain during the low-frequency enhancement conditions, particularly during the 24-dB SL condition. This was initially evident during biological listening checks of all stimulus conditions. To assess the situation, the same hearing aid test system described above was used to measure total harmonic distortion for all stimulus condition settings during playback. The same pure-tone sweep (200 Hz through 8000 Hz) again served as stimulus. The stimulus was routed through the digital equalizer, audiometer, and insert earphone while coupled to the 2-cm^3 HA-2 coupler. Sampling and measurement procedures were the same as described above. Initially, the system was calibrated so that the 1000-Hz calibration tone for the speech material on the compact disc peaked the VU meter on the audiometer at zero dB. Prior to running the pure-tone sweep through the system, the 1000-Hz pure tone of the sweep was equated in SPL with the 1000-Hz calibration tone in the 2-cm^3 HA-2 coupler. Confirming what was evident during the listening check, total harmonic distortion exceeded 30% at 600 Hz and below during the 24-dB SL low-frequency-enhanced condition. It was evident that excessive distortion was a result of saturation of the earphone output (i.e., SPL outputs in excess of 125 dB SPL) with excessive low-frequency energy peaking the audiometer VU meter. To alleviate this problem (i.e., minimize speech stimuli peaking the VU meter during low-frequency enhancement), the system was recalibrated so that the 1000-Hz calibration tone for the speech material on the compact disc peaked the VU meter on the audiometer at −20 dB. By doing so, total harmonic distortion was reduced to less than 5% for all playback conditions except the 24-dB SL low-frequency-enhanced condition at the output maximum of the audiometer (i.e., 107 dB SPL for speech stimuli). For that condition only, total harmonic distortion was approximately 20% at 600 Hz and below. At lower audiometric output (e.g., 102 dB SPL and below), total harmonic distortion was less than 5%. By reducing the audiometric maximum output by 20 dB, it restricted including participants to those that had ARTs no greater than 75 dB HL (i.e., 82 dB SPL). That is, the ART plus 24 dB SL was less than the audiometer maximum output of 107 dB SPL. Consequently, only participants with ARTs of 75 dB HL (i.e., 82 dB SPL) listened, unavoidably, to speech material that had total harmonic distortion that was greater than 5% and that happened for only one word list at the 24-dB SL low-frequency-enhanced condition.

3. Procedure

Prior to listening to the test stimuli, participants received pure-tone audiometry, tympanometry, and ipsilateral ART testing. All participants met an ART criterion of ⩽75 dB HL (82 dB SPL) to a broadband (125–4000 Hz) noise activator. A broadband noise activator was chosen to mimic the response of the stapedius reflex to speech material. The ear with the lowest ipsilateral ART was chosen as the test ear. In cases where the two ears had identical ARTs, one ear was chosen randomly. ARTs were determined with a clinical middle-ear analyzer (Grason Stadler, Inc., model GSI 33). ART was defined as a decrease in static admittance of ⩾0.02 cm^3. ART testing was carried out at the middle-ear pressure value where peak mobility occurred during tympanometry. A 5-dB step size was employed to determine the ART.

Participants were presented with the identical CVC stimuli at seven SLs *re*: their respective ARTs: −60, −48, −36, 0, 8, 16, and 24 dB. For the normal listening conditions, participants received the test stimuli at all seven SLs. For the simulated stapedius muscle paralysis listening (i.e.,

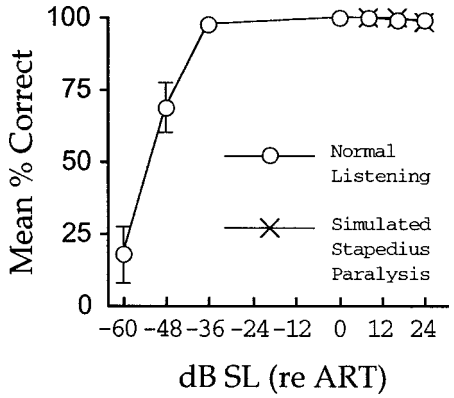


FIG. 2. Mean word recognition performance (i.e., percentage of 30 phonemes correct) as a function of listening condition (i.e., normal listening vs simulated stapedius muscle paralysis) and dB SL *re*: ART. Note that the low-frequency-enhanced stimuli were used only at the three highest stimulus levels, i.e., those in which the stapedius reflex would be expected to be evoked. Error bars represent plus/minus one standard error of the mean.

low-frequency enhancement), participants received the test stimuli only at 8, 16, and 24 dB SL *re*: ART for a total of ten listening conditions. At no time did the test stimuli exceed participant's loudness discomfort levels. Contralateral speech noise masking was applied when necessary (Martin *et al.*, 1991; Martin, 1997). The presentation of CVC lists was counterbalanced while the presentation order of the ten listening conditions was determined by a digram-balanced Latin square (Wagenaar, 1969). All CVC test stimuli were presented monaurally. A practice list was presented to participants prior to testing. Participants were required to provide a written response to the test stimuli.

B. Results and discussion

Participants' written responses were scored phonetically. Mean word recognition performance (i.e., percentage of 30 phonemes correct) as a function of listening condition (i.e., normal listening vs simulated stapedius muscle paralysis) and dB SL *re*: ART are shown in Fig. 2. As expected, participants exhibited a steep performance-intensity function for the normal listening condition. Participants displayed similar performance in the simulated stapedius muscle paralysis listening condition (i.e., low-frequency enhancement) to that of the normal listening condition at equivalent SLs.

The participants' proportional scores were transformed to ''rationalized'' arcsine units (Studebaker, 1985) prior to subjecting them to inferential statistical analyses. This simple linear arcsine transformation is recommended as ''proportional data are not well suited to inferential statistics because the means and the variances of such data are correlated'' (Studebaker, p. 461). Rationalized arcsine units (RAUs) are more attractive than arcsine units, as the former are numerically closer to the original percentage values while at the same time retain all of the desirable statistical properties of the latter. A two-factor repeated measures analysis of variance was undertaken to investigate mean word recognition performance as a function of listening condition (i.e., normal listening vs simulated stapedius muscle paralysis) and sensation level (8, 16, and 24-dB SL *re*: ART). Nonsignificant main effects were found for listening condition [$F(1,9) = 1.58$, Greenhouse–Geisser $p = 0.24$, $\omega^2 = 0.050$] and sensation level [$F(2,18) = 2.47$, Greenhouse–Geisser $p = 0.14$, $\omega^2 = 0.12$]. As well, the listening condition by sensation level interaction was nonsignificant [$F(2,18) = 2.13$, Greenhouse-Geisser $p = 0.17$, $\omega^2 = 0.097$]. In other words, word recognition performance was indistinguishable across both listening condition and the relevant sensation levels.

Experiment 1 revealed no effect on word recognition of offsetting the contribution of the stapedius reflex by low-frequency enhancement of the speech stimuli. The most obvious potential flaw in the design of experiment 1 is that the attempt to offset the high-pass filtering action of the stapedius by enhancing low-frequency stimulus content was itself offset by still stronger stapedial contractions driven by the more intense signal. Recall, however, that we used up to 20-dB enhancement of the low-frequency content in stimuli that were up to 24 dB above ART (for peak stimulus SPLs near 107 dB). The maximum attenuation provided by the stapedius reflex is in the order of 20 dB (see Wilson and Margolis, 1991). This suggests that there is little doubt that, at least at our highest stimulus levels, the high-pass filtering effect of the stapedius contraction was successfully offset by our low-frequency enhancement.

The maximum stimulus amplitude used in experiment 1 was similar to that used by Dorman *et al.* (1987; cf. 107 and 108 dB SPL, respectively), who found up to 50% rollover for brief synthetic vowels. In the present study, stimulus levels were set relative to the listeners' own ART (see above). For the listeners with the lowest ARTs, however, this had the consequence that the highest SPLs used were lower than those employed by Dorman *et al.* The failure to demonstrate rollover in the present study, however, is in all probability not due to weaker reflex activation secondary to the use of lower stimulus levels. This is because the present stimuli were, by design, set at levels relative to individual ARTs, and we were therefore confident that the reflex was activated. Studies of clinical patients which have reported significant (e.g., $\geqslant$50%) rollover typically have done so for stimulus levels comparable to those used here (e.g., above 100 dB SPL: Borg and Zakrisson, 1973; 95 and 110 dB SPL: Chadwell and Greenberg, 1979; 100 dB HL: Dirks *et al.*, 1977; approximately 100 dB HL: Gang, 1976; approximately 100 to 110 dB SPL: Hannley and Jerger, 1981; 100 dB HL: McCandless and Goering, 1974; 100 dB HL: Wormald *et al.*, 1995; 100–105 dB HL: Meyer and Mishler, 1985). The fact that the stimulus levels used by some of these earlier studies were modestly higher than ours might lead us to expect modestly weaker rollover in our study. However, rollover is a continuous, graded function of stimulus level, and one would therefore have expected that if rollover were weaker in our study, it would only be very modestly so. In practice, we saw virtually zero rollover in any of our normal listeners (Fig. 2).

The reasons for the discrepancy between our data on normal listeners and those of Dorman *et al.* (1987) are unclear, but one explanation is as follows. In the Dorman *et al.* task, the only basis available to the listeners for vowel identification was the strictly auditory analysis of the stimulus. This is because the stimuli they used had no semantic con-

TABLE I. Summary of preoperative and postoperative audiometric data for all participants with stapedectomy.

			Preoperative		Postoperative	
Participant	Age	Gender	Pure-tone average[a]	SRT	Pure-tone average[a]	SRT
1	45	Female	41.7	40	31.7	30
2	55	Female	26.7	25	13.3	15
3	53	Female	41.7	35	23.3	25
4	30	Female	53.3	55	21.7	20
5	43	Female	48.3	40	20.0	15
6	29	Male	50.0	45	26.7	30

[a]Pure-tone average represents three frequency average of thresholds at 500, 1000, and 2000 Hz. All thresholds are in dB HL.

tent. Moreover, their stimuli were exceptionally brief (i.e., 50 ms), so there was very little acoustic redundancy in the stimuli, and, accordingly, great fidelity was required in the auditory encoding process. In our task, however, some level of ambiguity in the output of the auditory encoding process could, in principle, have been offset by the listener's knowledge that each stimulus was a real word, whose identity could possibly be reconstructed from the conjunction of the listener's general knowledge and any unambiguously encoded parts of the acoustic signal. Our listeners thus had significant advantages over those in the Dorman *et al.* study. We shall return to the further ramifications of this argument in Sec. IV.

III. EXPERIMENT 2

The general purpose of experiment 2 was to provide an independent line of evidence on the contribution of the stapedius reflex to the rollover in word recognition at high stimulus intensities. Experiment 1 attempted to counteract the effects of the stapedius reflex in normal listeners by low-frequency enhancement of the speech stimuli. These listeners showed no evidence of rollover in their performance-intensity functions. In experiment 2, we have studied the performance-intensity functions of patients with surgically confirmed severing of the stapedius tendon.

A. Method

1. Participants

Six adults ($M=42.5$ years, s.d.$=11.1$; one male and five females) who had undergone a small fenestra stapedectomy served as participants. In all cases the incudostapedial joint was disarticulated and the stapedial tendon was cut. The superstructure of the stapes was removed. Following fenestration of the footplate, a graft was placed over the oval window. A Robinson prosthesis was placed with its bucket handle reflected over the long process of the incus, centered over the oval window, and then crimped in place. All but one of the participants was tested within one year following their stapedectomy in the test ear. Two of the six participants had undergone a previous stapedectomy in the nontest ear. Preoperative and postoperative audiometric data for all participants are presented in Table I. Participants also presented with normal middle-ear function as indexed by peak compensated static acoustic admittance, tympanometric width, and equivalent ear-canal volume (American Speech-Language-Hearing Association, 1997). The absence of an stapedius reflex was confirmed in the test ear with ipsilateral stapedius reflex testing.

2. Stimuli and apparatus

The test stimuli consisted of the same CVC word lists described in experiment 1. The same double-wall sound-treated audiometric suite described in the first experiment served as the test environment. The recorded stimuli were routed from the same compact disc player to the same clinical audiometer and presented to each participant through the insert earphone.

3. Procedure

Participants were presented with the identical CVC stimuli at 10 SLs (i.e., 10, 20, 30, 50, 60, 65, 70, 75, 80, and 85 dB) *re*: their respective SRTs. Contralateral speech-noise masking was applied when necessary (Martin *et al.*, 1991; Martin, 1997). The presentation of CVC lists was counterbalanced, while the presentation order of the ten listening conditions was determined by a digram-balanced Latin square (Wagenaar, 1969). All CVC test stimuli were presented monaurally. A practice list was presented to participants prior to testing. Participants were required to provide a written response to the test stimuli.

B. Results and discussion

Participants' written responses were scored phonetically. Word recognition performance (i.e., percentage of 30 phonemes correct) as a function of dB SL *re*: SRT are plotted separately for each of the six listeners in Fig. 3. As expected, participants exhibited steep, saturating performance-intensity functions. There was no evidence of any rollover in word recognition scores at the higher presentation levels for any of the participants, even at stimulus levels as high as 85 dB *re*: SRT (100–115 dB HL).

Experiment 2 described the performance-intensity functions for word recognition in quiet of patients with stapedectomy and normal or near-normal postoperative pure-tone average thresholds. The performance-intensity curves were saturating functions that showed no evidence of rollover. McCandless and Goering (1974) reported that some (approximately 35%) patients with stapedectomy studied for their speech audiometry showed significant rollover. These are important data, unless any coexisting cochlear or other

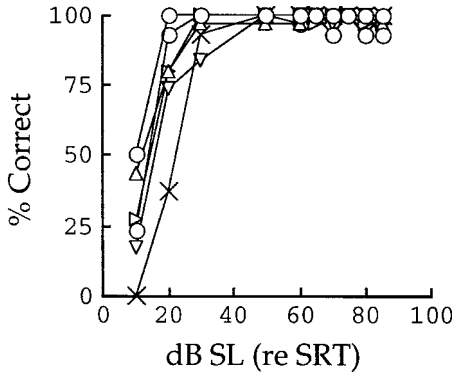


FIG. 3. Word recognition performance as a function of stimulus level (in dB *re*: SRT), plotted separately for each of the patients with stapedectomy.

hearing loss in that subset of patients confounds the finding. In this regard, the audiological data on some of these patients had significant high-frequency hearing losses (e.g., their Fig. 5; see also Gang, 1976). In the present experiment 2, none of the patients presented evidence of significant cochlear hearing loss, and none of them showed rollover in conventional speech audiometry with stimulus levels up to 85 dB SL *re*: SRT. Given that the patients' SRTs were in the range from 15 to 30 dB HL, this means that our maximum stimulus levels were in the range from 100 to 115 dB HL. These values are in the same range as those used by earlier clinical studies (e.g., above 100 dB SPL: Borg and Zakrisson, 1973; 95 and 110 dB SPL: Chadwell and Greenberg, 1979; 100 dB HL: Dirks *et al.*, 1977; approximately 100 dB HL: Gang, 1976; approximately 100 to 110 dB SPL: Hannley and Jerger, 1981; 100 dB HL: McCandless and Goering, 1974; 100 dB HL: Wormald *et al.*, 1995).

IV. GENERAL DISCUSSION

The purpose of this report has been to describe two experiments probing the role of the stapedius reflex in word recognition in quiet. It was prompted by the hypothesis that one role of the stapedius muscle is to prevent the masking of high-frequency content of speech by low-frequency content at high stimulus levels. An important line of evidence for this hypothesis historically came from the ''rollover'' seen in the word recognition performance of Bell's palsy patients; this, on the assumption that Bell's is an isolated pathology of the facial nerve. The veracity of that assumption is in doubt, because there is neurological evidence in some Bell's palsy patients of involvement of other cranial nerve and brainstem systems (Adour, 1975; Citron and Adour, 1978; Rosenthal *et al.*, 1983). In this regard, cochlear nerve or brainstem disease, with or without attendant stapedius reflex abnormalities, can be associated with significant rollover (Jerger and Jerger, 1971; Dirks *et al.*, 1977; Hannley and Jerger, 1981; Meyer and Mishler, 1985). This means that it is important to tease out whether the absence of the stapedius reflex alone is capable of causing rollover in word recognition.

Our experiment 1 found no evidence for an effect on word recognition at high stimulus levels of low-frequency enhancement designed to mimic loss of the stapedius reflex in normal listeners. Word recognition scores were close to 100% correct for all listeners, under both normal and low-frequency-enhanced conditions, even at the highest amplitudes employed. Those amplitudes were expressed in dB *re*: measured ART in those listeners, and we thus had confidence that the reflex was evoked by the stimuli. Moreover, the absolute stimulus levels were in the same range as those used in clinical reports of rollover in cases of facial nerve and brainstem disease. Our experiment 2 examined the performance-intensity functions of patients with surgically confirmed disconnection of the stapedius muscle. These listeners showed no rollover in their word recognition scores at high stimulus levels (up to 85 dB SL *re*: SRT). These two sets of data constitute independent, converging lines of evidence that the stapedius reflex has little or no effect on word recognition in quiet.

In experiment 1, the contrast with the findings of Dorman *et al.* (1987) led us to speculate that any role of stapedius contractions in the prevention of upward spread of masking is best revealed under listening conditions which place the most stringent demands on auditory encoding. Dorman *et al.* used very brief, isolated vowels that created this demand (as opposed to our own study using familiar whole words). The most obvious other such listening condition is speech in the presence of competing noise. In this regard, Liden *et al.* (1963), Chadwell and Greenberg (1979), and Colletti *et al.* (1988) all reported that patients with stapedectomy showed relatively poor speech recognition in noise; where comparisons were possible, these patients showed greater sensitivity to noise masking than did patients with stapedectomy with preserved stapedius tendons (Liden *et al.*, 1963; Colletti *et al.*, 1988).

On the other hand, Middelweerd *et al.* (1989) found no significant threshold shift in signal-to-noise ratio for the intelligibility of sentences presented in noise among 37 consecutive patients with stapedectomy. They also pointed out that many of the patients in the earlier study by Chadwell and Greenberg (1979) had high-frequency hearing losses (4.0 kHz and above), and that this might have been a confounding factor in that study. Certainly, the fact that listeners with high-frequency hearing losses have impoverished recognition of speech in noise, especially when the noise has a fluctuating amplitude envelope [as might be expected in the case of low-frequency maskers of the kind used by Chadwell and Greenberg (1979) and by Liden *et al.* (1963)], has received recent experimental support (Phillips *et al.*, 1994; Stuart and Phillips, 1996). Detailed pure-tone threshold data were not presented for the patients with stapedectomy studied by Liden *et al.* (1963) or Colletti *et al.* (1988), and so it is unknown whether the Middelweerd *et al.* (1989) argument is relevant to those studies. Other recent reports have presented evidence on auditory sensitivity after stapedectomy or stapedotomy, but have not provided detailed information on listeners' speech audiometry, with or without noise masking (Ginsberg *et al.*, 1981; Levy *et al.*, 1990; Langman *et al.*, 1991). It seems, then, that speech-in-noise studies of patients with stapedectomy and stapedotomy have not yet provided clear support for the protection from upward spread of masking hypothesis of stapedius function. The case is arguably made still worse by Chadwell and Greenberg's (1979) obser-

vation that the particular consonant recognition errors made by patients with stapedectomy under conditions of noise masking were not necessarily those predicted by an upward spread of masking hypothesis. We are not aware of any detailed reports of performance-intensity functions of patients with stapedectomy for speech in noise, let alone in patients with acceptable postoperative audiograms.

None of the foregoing necessarily disputes the physiological findings in normal ears that stapedius muscle contractions can effect large ''unmasking'' effects of low-frequency noise on the cochlear neural responses to high-frequency signals [''large,'' because the (linear, dB/dB) effect of stapedius contractions on signal transmission might be enhanced in the cochlea by a rapid rate of growth of masking (e.g., 2 dB/dB; see Pang and Guinan, 1997)]. Our concern is that the available evidence on the strictly behavioral expression of this effect appears to be restricted to the most adverse listening conditions, e.g., in the recognition of isolated, very brief vowels (Dorman *et al.*, 1987) or perhaps speech in noise (Liden *et al.*, 1963; Colletti *et al.*, 1988). Certainly, the present data provided no evidence for a role of the stapedius reflex in word recognition in quiet.

ACKNOWLEDGMENTS

Preparation of this report was supported in part by grants from NSERC of Canada to DPP. Special thanks are due to Dr. Walter Green for his very helpful discussions at all stages of this work. The generous assistance of P. Bradley Brechtelsbauer, M. D., East Carolina ENT—Head and Neck Surgery, Greenville, NC in the recruitment of the participants with stapedectomy is greatly appreciated.

Adour, K. K. (**1975**). ''The bell tools for decompression,'' N. Engl. J. Med. **292**, 748–750.

ANSI (**1996**). ANSI S3.6-1996, ''Specification for audiometers'' (American National Standards Institute, New York).

ANSI (**1999**). ANSI S3.1-1999, ''Permissible ambient noise levels for audiometric test rooms'' (American National Standards Institute, New York).

American Speech-Language-Hearing Association (**1997**). *Guidelines for Audiological Screening* (American Speech-Language-Hearing Association, Rockville Pike, MD).

Ardic, F. N., Topalogu, I., Öncel, S., Ardic, F., and Uguz, M. Z. (**1997**). ''Does the stapes reflex remain the same after Bell's palsy?'' Am. J. Otol. **18**, 761–765.

Borg, E. (**1968**). ''A quantitative study of the effect of the acoustic stapedius reflex on sound transmission through the middle ear of man,'' Acta Oto-Laryngol. **66**, 461–472.

Borg, E., and Counter, S. A. (**1989**). ''The middle-ear muscles,'' Sci. Am. **261**, 74–80.

Borg, E., and Zakrisson, J. E. (**1973**). ''Stapedius reflex and speech features,'' J. Acoust. Soc. Am. **54**, 525–527.

Borg, E., and Zakrisson, J. E. (**1974**). ''Stapedius reflex and monaural masking,'' Acta Oto-Laryngol. **78**, 155–161.

Borg, E., and Zakrisson, J. E. (**1975**). ''The activity of the stapedius muscle in man during vocalization,'' Acta Oto-Laryngol. **79**, 325–333.

Brigham Young University (**1991**). Speech Audiometry Material [CD]. Provo, UT: Hearing and Speech Sciences Laboratory, Brigham Young University.

Chadwell, D. L., and Greenberg, H. J. (**1979**). ''Speech intelligibility in stapedectomized individuals,'' Am. J. Otol. **1**, 103–108.

Citron, D. III, and Adour, K. K. (**1978**). ''Acoustic reflex and loudness discomfort in acute facial paralysis,'' Arch. Otolaryngol. **104**, 303–306.

Colletti, V., Sittoni, V., and Fiorino, F. G. (**1988**). ''Stapedotomy with and without stapedius tendon preservation versus stapedectomy: Long-term results,'' Am. J. Otol. **9**, 136–141.

Dirks, D. D., Kamm, C., Bower, D., and Betsworth, A. (**1977**). ''Use of performance-intensity functions for diagnosis,'' J. Speech Hear Disord. **42**, 408–415.

Dorman, M. F., Cedar, I., Hannley, M. T., Lindholm, J. M., and Leek, M. (**1986**). ''The role of the acoustic reflex in maintaining vowel intelligibility at high sound pressure levels,'' J. Speech Hear. Res. **29**, 420–424.

Dorman, M. F., Lindholm, J. M., Hannley, M. T., and Leek, M. R. (**1987**). ''Vowel intelligibility in the absence of the acoustic reflex: Performance-intensity characteristics,'' J. Acoust. Soc. Am. **81**, 562–564.

Gang, R. P. (**1976**). ''The effects of age on the diagnostic utility of the rollover phenomenon,'' J. Speech Hear Disord. **41**, 63–69.

Ginsberg, I. A., Hoffman, S. R., White, T. P., and Stinziano, G. D. (**1981**). ''Hearing changes following stapedectomy: A six year follow-up,'' Laryngoscope **91**, 87–92.

Hannley, M., and Jerger, J. (**1981**). ''PB rollover and the acoustic reflex,'' Audiology **20**, 251–258.

Hendrix, R. A., and Melnick, W. (**1983**). ''Auditory brain stem response and audiologic tests in idiopathic facial nerve paralysis,'' Otolaryngol.-Head Neck Surg. **91**, 686–690.

Jerger, J., and Jerger, S. (**1971**). ''Diagnostic significance of PB word functions,'' Arch. Otolaryngol. **93**, 573–580.

Langman, A. W., Jackle, R. K., and Sooy, F. A. (**1991**). ''Stapedectomy: Long-term hearing results,'' Laryngoscope **101**, 810–814.

Levy, R., Shvero, J., and Hadar, T. (**1990**). ''Stapedotomy technique and results: Ten years' experience and comparative study with stapedectomy,'' Laryngoscope **100**, 1097–1099.

Liden, G., Nordlund, B., and Hawkins, J. E., Jr. (**1963**). ''Significance of the stapedius reflex for the understanding of speech,'' Acta Oto-Laryngol., Suppl. **188**, 275–279.

Lundgren, A., Ödkvist, L. M., Hendriksson, K. G., Larsson, L. E., Karlberg, B. E., and Jerlvall, L. (**1977**). ''Facial palsy in diabetes meillitus—not only a mononeuropathy?'' Adv. Oto-Rhino-Laryngol. **22**, 182–189.

Mahoney, T., Vernon, J., and Meikle, M. (**1979**). ''Function of the acoustic reflex in discrimination of intense speech,'' Arch. Otolaryngol. **105**, 119–123.

Martin, F. N. (**1997**). *Introduction to Audiology*, 6th ed. (Allyn and Bacon, Boston).

Martin, F. N., Severagnce, G. K., and Thibodeau, L. (**1991**). ''Insert earphones for speech audiometry,'' J. Am. Acad. Audiol **2**, 55–58.

Maurizi, M., Ottaviani, F., Almadori, G., Falchi, M., and Paludetti, G. (**1987**). ''Auditory brainstem and middle-latency responses in Bell's palsy,'' Audiology **26**, 111–116.

McCandless, G. A., and Goering, D. M. (**1974**). ''Changes in loudness after stapedectomy,'' Arch. Otolaryngol. **100**, 344–350.

McCandless, G. A., and Schumacher, M. H. (**1979**). ''Auditory dysfunction with facial paralysis,'' Arch. Otolaryngol. **105**, 271–274.

Meyer, D. H., and Mishler, E. T. (**1985**). ''Rollover measurements with Auditec NU-6 word lists,'' J. Speech Hear Disord. **50**, 356–360.

Middelweerd, M. J., Feenstra, L., van der Baan, S., and Plomp, R. (**1989**). ''The effect of stapedectomy on speech intelligibility in noise,'' Am. J. Otol. **5**, 380–384.

Nuttall, A. L. (**1974**). ''Tympanic muscle effects on middle-ear transfer characteristic,'' J. Acoust. Soc. Am. **56**, 1239–1247.

Olsen, W. O., Van Tasell, D. J., and Speaks, C. E. (**1997**). ''Phoneme and word recognition for words in isolation and in sentences,'' Ear Hear. **18**, 175–188.

Pang, X. D., and Guinan, J. J., Jr. (**1997**). ''Effects of stapedius-muscle contractions on the masking of auditory-nerve responses,'' J. Acoust. Soc. Am. **102**, 3576–3586.

Phillips, D. P., and Carr, M. M. (**1998**). ''Disturbances of loudness perception,'' J. Am. Acad. Audiol **9**, 371–379.

Phillips, D. P., Rappaport, J. M., and Gulliver, J. M. (**1994**). ''Impaired intelligibility of speech in noise in patients with noise-induced cochlear hearing loss: Contribution of temporal processing defect,'' Am. J. Otol. **15**, 679–686.

Rabinowitz, W. M. (**1977**). ''Acoustic-reflex effects on the input admittance and transfer characteristics of the human middle ear,'' Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge.

Rosenthal, U., Edström, S., Hanner, P., and Badr, G. (**1983**). ''Auditory brain stem response abnormalities in patients with Bell's palsy,'' Otolaryngol.-Head Neck Surg. **91**, 412–416.

Stevens, S. S., and Davis, H. (**1938**). *Hearing* (Wiley, New York).

Stuart, A., and Phillips, D. P. (**1996**). ''Word recognition in continuous and

interrupted broadband noise by young normal-hearing, older normal hearing, and presbyacusic listeners,'' Ear Hear. **17**, 478–489.

Studebaker, G. (**1985**). ''A rationalized arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.

Wagenaar, W. A. (**1969**). ''Note on the construction of digram-balanced Latin squares,'' Psychol. Bull. **72**, 384–386.

Wilson, R. H., and Margolis, R. H. (**1991**). ''Acoustic-reflex measurements,'' in *Hearing Assessment*, 2nd ed., edited by W. F. Rintelmann (Pro-Ed, Austin), pp. 247–319.

Wormald, P. J., Rogers, C., and Gatehouse, S. (**1995**). ''Word recognition in patients with Bell's palsy and a paralyzed stapedius muscle,'' Clin. Otolaryngol. **20**, 59–62.

Zakrisson, J. E., Borg, E., and Blom, S. (**1974**). ''The acoustic impedance change as a measure of stapedius muscle activity in man,'' Acta Oto-Laryngol. **78**, 357–364.

Cooperative interaction as the physical basis of the negative stiffness in hair cell stereocilia

K. H. Iwasa[a)]
Section on Biophysics, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, Maryland 20892

G. Ehrenstein
Biophysics Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892

(Received 26 July 2001; accepted for publication 8 February 2002)

A recent report confirmed that stiffness of the stereocilia can be negative, as predicted by the Howard–Hudspeth model. According to this model, the mechanotransducer channel's gating not only reduces the stereociliary stiffness, but can alter its sign as well. The basic assumptions of this model do not include cooperativity in channel gating. Here we consider two possible explanations for the observed negative stiffness. If the stereocilia have a special structure so that microscopic displacement can be imposed on each channel by controlling the bending of the bundle, negative stiffness can occur without channel cooperativity. If such a microscopic condition cannot be imposed by a macroscopic manipulation, an additional physical process, such as cooperativity in channel gating, is required to explain negative stiffness. [DOI: 10.1121/1.1466864]

PACS numbers: 43.64.Ld, 43.64.Kc, 43.64.Bt [LHC]

I. INTRODUCTION

Measurement of ''gating compliance,'' an increase in the compliance of the stereocilia associated with gating of the mechanotransducer channel, has demonstrated that this channel can be gated by mechanical force (Howard and Hudspeth, 1987, 1988). To describe this phenomenon quantitatively, Howard and Hudspeth (1988) introduced a two-state model, which was successful not only in explaining ''gating stiffness,'' but also in predicting negative stiffness (Martin *et al.*, 2000) associated with movement of the stereocilia.

The significance of negative stiffness in mechanosensory systems is that hair bundles not only sense force passively, but also produce force in response to applied force. Such active force can reduce or even eliminate the effect of viscous damping resulting from the aqueous environment of hair bundles. Thus this mechanism insures both sensitivity and frequency specificity of the system (Eguíluz *et al.*, 2000; Manley, 2000).

To clarify the physical basis of the negative stiffness of hair bundles, we first compare the Howard–Hudspeth model with the Sachs–Lecar model (Sachs and Lecar, 1991), an alternative approach that also leads to ''gating compliance.'' We then discuss whether and how cooperative interactions can be included.

Although the Sachs–Lecar model was proposed for stretch-activated channels, which are mechanosensitive channels distributed in a wide range of cells, it can be transparently compared with the Howard–Hudspeth model. Both models assume that mechanosensitive channels undergo conformational transitions that include changes in the size of the channels. The Howard–Hudspeth paper assumes that the system, which consists of a tip-link and the channel, is one-dimensional. It also assumes that the tip-link has a finite stiffness and that the channel is infinitely stiff. For this reason, the tip-link gives the characteristic stiffness of the system, independent of the conformation of the channel. The Sachs–Lecar paper assumes that a stretch-activated channel is a two-dimensional object that has its own finite stiffness in an open conformation and in a closed conformation. The stiffness of these conformations may not be the same. If membrane tension is isotropic, this system can be treated one dimensionally. If membrane tension or the compliance of the channel is anisotropic and is larger in one direction, it can also treated one-dimensionally.

The treatment by Howard–Hudspeth and that by Sachs–Lecar have been compared previously (Lecar and Morris, 1993; Corey and Howard, 1994). An obvious and well-appreciated difference between the two treatments is whether or not a stiffness difference between open and closed conformations of a mechanosensitive channel is important. In two previous reports (Lecar and Morris, 1993; Corey and Howard, 1994), the terms ''Howard–Hudspeth model'' and ''Sachs–Lecar model'' were used to indicate this difference. Here, we do not use these terms in the same way. Our assumption on channel stiffness is the same as the Howard–Hudspeth paper in not considering the stiffness difference. A recent study of stretch-activated channels shows that the channel undergoes clear size changes but stiffness changes are more subtle (Sukharev *et al.*, 1999). We use the terms ''Howard–Hudspeth model'' and ''Sachs–Lecar model'' to distinguish the *statistical models* used to derive the stress–strain relationship of mechanosensitive channels in the two papers.

We treat a mechanotransducer channel as a one-dimensional object. We further assume that the channel has

a)Electronic mail: iwasa@nih.gov

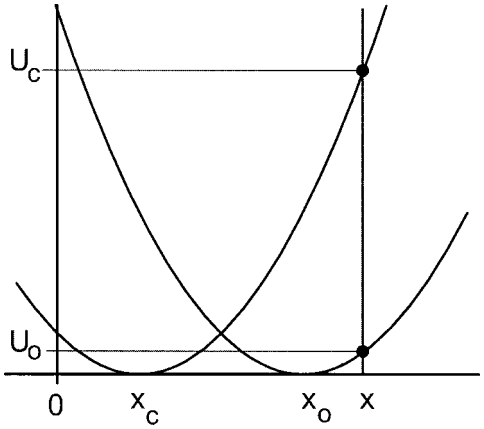


FIG. 1. Internal energy diagrams for closed and open configurations of a transducer channel. For a given length of the channel, energy difference ΔU of the conformations is given by $\Delta U = -k(x_o - x_c)x + \Delta U_0$. A special case of $\Delta U_0 = 0$ is illustrated.

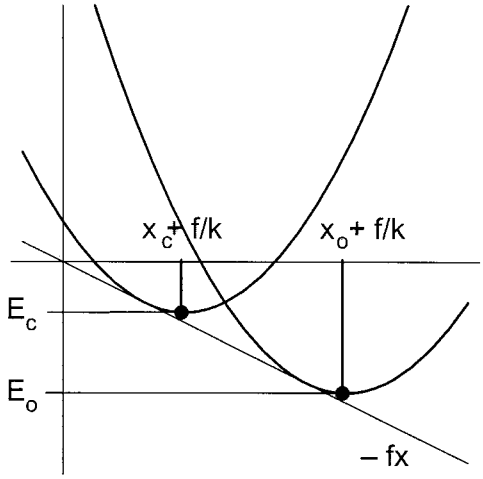


FIG. 2. The free energy profiles of closed and open configurations. Stretching force f favors the open state by reducing its energy level ΔE relative to the closed state. The free energy difference ΔE is given by $\Delta E = -(x_o - x_c)f + \Delta E_0$. A special case of $\Delta E_0 = 0$ is illustrated.

two states that differ in the resting length and that the channel is associated with a spring of its own or of the tip-link, the stiffness of which does not depend on the state of the channel. Based on these simplifying assumptions, we show that the two models are based on different statistical models relating macroscopic and microscopic variables of the system under consideration. As a result, although both models predict an increase in the compliance of the mechanotransducer system under the condition that gating occurs, they do not yield the same functional dependence. In particular, the Sachs–Lecar model predicts negative compliance only when cooperativity is present, whereas the Howard–Hudspeth model predicts negative compliance whether or not cooperativity is present.

II. HOWARD–HUDSPETH MODEL

Howard and Hudspeth (1988) gave a framework for quantitatively describing the gating of mechanotransducer channels. Their model was further examined by Markin and Hudspeth (1995). They assume that the closed state and the open state of the transducer channel have different resting lengths and that displacements of the channel results in displacing a spring with stiffness k (Fig. 1). They then impose the same length x to both states. The elastic energy $U_c(x)$ of the closed state and the energy $U_o(x)$ of the open state are represented by $U_c(x) = \frac{1}{2}k(x - x_c)^2 + U_{c0}$ and $U_o(x) = \frac{1}{2}k(x - x_o)^2 + U_{o0}$. The difference ΔU of the energy in the two states is thus

$$\Delta U(x) = U_o(x) - U_c(x) = -k\Delta x \cdot x + \Delta U_0, \tag{1}$$

where ΔU_0 is a constant and $\Delta x = x_o - x_c$.

The open probability P_o of the channel is given by

$$P_o(x) = \frac{\exp[-\beta \Delta U(x)]}{1 + \exp[-\beta \Delta U(x)]}, \tag{2}$$

where $\beta = 1/(k_B T)$ with Boltzmann's constant k_B and the temperature T.

The mean force $\langle f(x) \rangle$ that is felt by the channel when its length is x is given by

$$\begin{aligned}\langle f(x) \rangle &= k(x - x_c)(1 - P_o(x)) + k(x - x_o)P_o(x) \\ &= k(x - x_c) - k\Delta x P_o(x).\end{aligned} \tag{3}$$

Thus the stiffness of the system is

$$\frac{d\langle f(x) \rangle}{dx} = k - \beta k^2 \Delta x^2 P_o(x)(1 - P_o(x)). \tag{4}$$

The first term reflects the stiffness of the system without gating. The second term is due to gating of the channel. This term not only reduces the stiffness, but it can change the sign of the stiffness as well. Since the maximum value of $P_o(1 - P_o)$ is $\frac{1}{4}$, the condition for negative stiffness is $k(x_o - x_c)^2 > 4k_B T$.

III. SACHS-LECAR MODEL

Sachs and Lecar (1991) proposed a model for describing stretch-activated channels, which also serves as an alternative approach for describing the mechanotransducer channel (Lecar and Morris, 1993; Corey and Howard, 1994). In the Sachs–Lecar model, the free energy difference of the two conformations depends on applied force f because mechanical work $f\,\Delta x$ is needed to close the channel. Both closed and open states rest in their respective energy minimum, the difference of which is determined by $\Delta E = -f\Delta x + \Delta E_0$ (Fig. 2). Then open probability P_o is expressed by

$$P_o(f) = \frac{\exp[-\beta \Delta E(f)]}{1 + \exp[-\beta \Delta E(f)]}, \tag{5}$$

where ΔE_0 is a constant. The mean length $\langle x \rangle$ of the channel is given by

$$\begin{aligned}\langle x \rangle &= \left(x_c + \frac{f}{k}\right)(1 - P_o(f)) + \left(x_o + \frac{f}{k}\right)P_o(f) \\ &= x_c + \frac{f}{k} + \Delta x P_o(f).\end{aligned} \tag{6}$$

Equations (3) and (6) appear similar if we disregard the distinction between x and $\langle x\rangle$. The functional dependence of P_o is different in the two equations.

The compliance of the system is then obtained by differentiating the mean length $\langle x\rangle$ with respect to force F,

$$\frac{d\langle x\rangle}{df}=\frac{1}{k}+\beta\Delta x^2 P_o(f)(1-P_o(f)). \tag{7}$$

Both terms in the compliance are positive. Gating increases the compliance. The compliance expressed by (7), however, cannot be negative.

IV. COMPARISON OF THE TWO MODELS

Here we notice that the Howard–Hudspeth model and the Sachs–Lecar model give different results. Both models predict a reduction of the stiffness or, equivalently, an increase of the compliance as the result of channel gating. Although the functional dependencies of these quantities are not the same, predictions of these models are identical when the gating term in Eq. (4) or (7) is small. The difference is striking, however, when the effect of gating is relatively large. For the condition in which the Howard–Hudspeth model predicts negative stiffness, the Sachs–Lecar model predicts positive compliance.

The difference in the predictions of the two models stems from the distinction between a mean length $\langle x\rangle$ and an unaveraged length x. The Howard–Hudspeth model assumes that it is possible to impose a length x on each channel. For example, it assumes that by controlling the macroscopic displacement of a stereocilia bundle, the microscopic displacement of each channel is controlled. This imposes the condition that both closed and open states have the same length x. Thus the open probability is determined by the energy difference ΔU between the open state with length x and the closed state with length x (Fig. 1). In most systems, imposing a microscopic condition by macroscopic means is not possible. Rather, the imposition of a macroscopic length on a microscopic system controls not the precise x, but the mean value $\langle x\rangle$, which is determined as a weighted average of the channel in the two states.

There are a number of necessary conditions for the Howard–Hudspeth model to be experimentally relevant. One such condition is that the transducer channels are connected functionally in parallel. Another condition is that the tip links and the transducer channels are fine tuned so that an equal displacement takes place on each channel for a given displacement applied to the bundle.

Mean length $\langle x\rangle$ in the Sachs–Lecar model appears as a mean of length in the closed state and length in the open state. This model assumes that there is no way of imposing a given length to each channel, but that imposing a mean length is always possible. That assumption is consistent with most systems. It is certainly true for stretch-activated channels and membrane based motors (Iwasa, 1994). With this assumption, negative stiffness does not appear in a two-state model without consideration of additional physical processes, such as cooperativity.

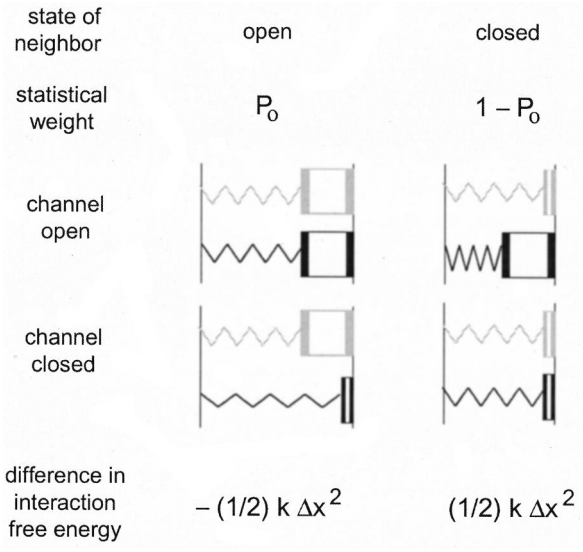


FIG. 3. The effect of nearest-neighbor interaction when channels are connected in parallel. The probability that the nearest neighbor is open is P_o and the probability that it is closed is $1-P_o$. The state of the channel is represented in black and the neighbor in gray. If the channel and the nearest neighbor are both closed or open, the interaction energy is zero. When one is open and the other is closed, the channel is either pulled or pushed by the neighbor. The interaction energy is $(\frac{1}{2})k\Delta x^2$ because it is equivalent of pulling or pushing an object that has stiffness k by the distance Δx. The interaction free energy is represented by $-P_0\cdot(\frac{1}{2})k\Delta x^2+(1-P_o)\cdot(\frac{1}{2})k\Delta x^2$.

In the framework of the Sachs–Lecar model, negative stiffness can be obtained as the result of cooperativity in channel gating. A simple way of introducing cooperativity is to assume that the free energy difference between the two states is dependent on the open probability of channels. For example, we can write

$$\Delta E=-f\Delta x-\gamma P_o+\Delta E_0, \tag{8}$$

where γ represents interaction with neighboring channels. With the assumption (8), the open probability P_o of the channel is given by

$$\frac{P_o}{1-P_o}=\exp[\beta(f\Delta x+\gamma P_o-\Delta E_0)]. \tag{9}$$

Equation (9) shows that an increase in the open probability P_o increases the right-hand side through the second term. An increase in the right-hand side in turn requires an increase in P_o. Thus the γP_o is a typical cooperativity term in the mean field approximation.

A positive value for γ implies positive cooperativity and a negative value negative cooperativity. In general, at least an additional parameter is required to describe cooperativity. It is, however, possible to make specific assumptions to determine γ explicitly without an additional parameter. Here we assume that a particular channel is in parallel connection with a neighboring channel (Fig. 3). Parallel connection imposes the length of the neighbor on the channel. The neighbor is either open or closed. We can control only mean values averaged over time. The statistical weight of the neighboring channel being open is P_o, which is the same as the channel. If the channel and the neighbor are not in the

same conformation, the neighbor pulls or pushes the particular channel. The energy due to pulling and pushing is $\frac{1}{2}k\Delta x^2$ because it is equivalent of displacing the channel (or the tip link), which has stiffness k, by the distance of Δx, which is from one state to the other.

Now we want to calculate the difference in interaction energy between the open and closed states of the particular channel. It is convenient to break up this interaction energy difference into two components—one component for the case where the neighboring channel is open and the other component for the case where the neighboring channel is closed. As represented diagrammatically in Fig. 3, the probability that the neighboring channel is open is P_o and the energy difference for this case is:

$$0-\tfrac{1}{2}k\Delta x^2.$$

The probability that the neighboring channel is closed is $(1-P_o)$, and the energy difference for this case is

$$\tfrac{1}{2}k\Delta x^2-0.$$

Thus, the contribution ΔE_{para} of parallel connectivity to the free energy difference ΔE can be expressed by (see Fig. 3)

$$\Delta E_{\text{para}}=P_o(0-\tfrac{1}{2}k\Delta x^2)+(1-P_o)(\tfrac{1}{2}k\Delta x^2-0) \tag{10}$$

$$=\tfrac{1}{2}k\Delta x^2-k\Delta x^2P_o\,. \tag{11}$$

Because the first term is a constant and does not affect the functional dependence of ΔE_{para}, it can be included in the constant term ΔE_0. Thus, this model is equivalent to putting

$$\gamma=k\Delta x^2, \tag{12}$$

which shows parallel connectivity results in positive cooperativity.

For simplicity, we illustrated only one immediate neighbor in Fig. 3. It is easy to see that the existence of two or more immediate neighbors does not change the result.

It turns out that Eq. (8) with (12) leads to the expression (4) for the stiffness, which is obtained by the Howard–Hudspeth model. Under this assumption, the free energy difference ΔE is a function of $\langle x\rangle$. Specifically, $\Delta E=-k\Delta x(\langle x\rangle-x_c)+\Delta E_0$. Then the stiffness $df/d\langle x\rangle$ of the system can be obtained by differentiating both sides of (6) with respect to $\langle x\rangle$, treating P_o as a function of $\langle x\rangle$. The result is the same as the expression (4) for stiffness in the Howard–Hudspeth model.

Although we treated the force f applied to the channel as unaveraged, indeed we are able to apply an ensemble averaged force $\langle f\rangle$ rather than unaveraged force f to individual channels without changing our results. Thus we can replace f by $\langle f\rangle$ in the above equations.

Gating of mechanoreceptor channels could be also mediated by certain ions or other intracellular messengers. Of these potential messengers, Ca ions contribute to *negative*, rather than *positive*, cooperativity between the channels (Howard and Hudspeth, 1988; Ricci *et al.*, 2000; Hudspeth *et al.*, 2000).

One other factor that can affect the stiffness is the adaptation motor. On increase of Ca ions, the adaptation motor moves to reduce tension applied to tip-links (Howard and Hudspeth, 1987; Hacohen *et al.*, 1989). This effect reduces the apparent stiffness of the stereocilia, but it is slower than the gating of the transducer channel. It has the effect of changing the operation point of the channels to ensure that the system remains sensitive, rather than affecting channel gating directly (Howard and Hudspeth, 1987; Hacohen *et al.*, 1989).

The descriptions of a mechanosensitive channel given earlier nominally consider a single channel. How are the results derived by such a condition related to a system of N channels? Specifically, what results are expected for series and parallel connections?

The average used in the derivations can mean either time average or an ensemble average, which are equivalent in statistical mechanics. If the channels are connected in parallel, the macroscopic force F and macroscopic displacement X are given by $F=N\langle f\rangle$ and $X=\langle x\rangle$. If they are connected in series, $F=\langle f\rangle$ and $X=N\langle x\rangle$.

It has been informally argued that parallel connection produces cooperativity because if one channel gates, the force applied to the rest of the channels increases for a given macroscopic force F, and that this constitutes cooperative interaction. We have shown that such an effect can be indeed incorporated into the Sachs–Lecar model, giving rise to the Howard–Hudspeth equation. However, such an argument on cooperativity does not apply to the Howard–Hudspeth model. In this model, force F applied to a system of N channels connected in parallel would be expressed by

$$F=N[f_c(1-P_o(x))+f_oP_o(x)], \tag{13}$$

where f_c and f_o are, respectively, the force applied to a closed channel and that applied to an open channel, and the open probability of the channels $P_o(x)$ for displacement x $(=X)$ is given by Eq. (2). The number of closed channels is expressed by $N(1-P_o)$ and the number of open channels is NP_o. If it were true that parallel connectivity produces cooperativity, Eq. (13) should include cooperativity. However, it is quite obvious that Eq. (13) is equivalent to Eq. (3), because the mean force per channel is given by $f=F/N$, and we have $f_c=k(x-x_c)$ and $f_o=k(x-x_o)$. Since the Howard–Hudspeth derivation considers only one channel and does not consider the state of neighboring channels, it does not include cooperativity. Thus, the equivalence of Eqs. (2) and (13) indicates that cooperative interactions are not involved.

In the derivation of a cooperative term (12), we considered parallel connection between a particular channel of interest and a neighboring channel. We assumed that the neighboring channel has a time-averaged open probability P_o. The interaction term (12) can be derived by assuming that many identical channels are connected in parallel. The displacements imposed on the channel of interest by the neighboring channels is ΔxP_o (from the channel's closed position) because P_o is the ensemble-averaged open probability of neighboring channels. Fluctuation of the imposed displacement may be small because of the large member of channels. In the Howard–Hudspeth derivation, imposed displacement has no fluctuation because they assumed that microscopic displacement can be imposed. Thus, imposing a microscopic

displacement has the same effect as introducing cooperativity between channels, and both approaches lead to the same result.

Negative stiffness appears in physics literature in conPo. The interaction term (12) can be derived by assuming that many identical channels are connected in parallel. The displacement imposed on the channel of interest by the neighboring channels is $\delta x p_o$ (from the channel's closed position) because P_o is the ensemble-averaged open probability of neighboring channels.nection with buckling of elastic objects (Bazant and Cedolin, 1991) and ferroelasticity (Knowles and Steinberg, 1978; Lakes, 2001). The former is deterministic and is related to deformations of macroscopic objects. Buckling increases compliance with increased strain, resulting in negative stiffness. In contrast, the ferroelastic form of negative stiffness is statistical and involves cooperative interactions between microscopic objects analogous to ferroelectricity and ferromagnetism (Salje, 1990).

Negative stiffness resulting from Sachs–Lecar model (Sachs and Lecar, 1991) is analogous to ferroelasticity (Salje, 1990) in that it is the consequence of cooperative interactions.

V. CONCLUSION

There are two possible mechanisms for negative stiffness of the mechanotransducer system being observed:

One possibility is an exquisitely fine-tuned organization of the stereocilia bundle, making it possible to impose a one-to-one relationship between the bending of the bundle and the displacement of each transducer channel.

Another possibility is that such a stringent condition for channel displacements cannot be realized by a macroscopic manipulation, and that the observed negative stiffness is a demonstration that gating of the mechanotransducer channel involves cooperativity.

ACKNOWLEDGMENTS

We thank Dr. Jim Hudspeth and anonymous reviewers for useful comments. We also thank Dr. David Ehrenstein for informing us of literature on ferroelasticity.

Bazant, Z., and Cedolin, L. (**1991**). *Stability of Structures* (Oxford U. P., Oxford, UK).

Corey, D. P., and Howard, J. (**1994**). ''Models for ion channel gating with compliant states,'' Biophys. J. **66**, 1254–1257.

Eguíluz, V. M., Ospeck, M., Choe, Y., Hudspeth, A. J., and Magnasco, M. O. (**2000**). ''Essential nonlinearities in hearing,'' Phys. Rev. Lett. **84**, 5232–5235.

Hacohen, N., Assad, J. A., Smith, W. J., and Corey, D. P. (**1989**). ''Regulation of tension on hair-cell transduction channels: displacement and calcium dependence,'' J. Neurosci. **9**, 3988–3997.

Howard, J., and Hudspeth, A. J. (**1987**). ''Mechanical relaxation of the hair bundle mediates adaptation in mechanoelectrical transduction by the bullfrog's saccular hair cell,'' Proc. Natl. Acad. Sci. U.S.A. **84**, 3064–3068.

Howard, J., and Hudspeth, A. J. (**1988**). ''Compliance of the hair bundle associated with gating of mechanoelectrical transduction channels in the bullfrog's saccular hair cell,'' Neuron **1**, 189–199.

Hudspeth, A. J., Choe, Y., Mehta, A. D., and Martin, P. (**2000**). ''Putting ion channels to work: mechanoelectrical transduction, adaptation, and amplification by hair cells,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 11765–11772.

Iwasa, K. H. (**1994**). ''A membrane model for the fast motility of the outer hair cell,'' J. Acoust. Soc. Am. **96**, 2216–2224.

Knowles, J. K., and Sternberg, E. (**1978**). ''On the failure of ellipticity and the emergence of discontinuous deformation gradients in the plane finite elastostatics,'' J. Elast. **8**, 329–379.

Lakes, R. S. (**2001**). ''Extreme damping in composite materials with a negative stiffness phase,'' Phys. Rev. Lett. **86**, 2897–2900.

Lecar, H., and Morris, C. E. (**1993**). ''Biophysics of mechanotransduction,'' in *Mechnoreception by the Vascular Wall*, edited by G. M. Rubanyi (Futura, Mount Kisco, NY), pp. 1–11.

Manley, G. A. (**2000**). ''Cochlear mechanisms from a phylogenetic viewpoint,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 11736–11743.

Markin, V. S., and Hudspeth, A. J. (**1995**). ''Gating-spring models of mechanoelectrical transduction by hair cells of the internal ear,'' Annu. Rev. Biophys. Biomol. Struct. **24**, 59–83.

Martin, P., Mehta, A. D., and Hudspeth, A. J. (**2000**). ''Negative hair-bundle stiffness betrays a mechanism for mechanical amplification by the hair cell,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 12026–12031.

Ricci, A. J., Crawford, A. C., and Fettiplace, R. (**2000**). ''Active hair bundle motion linked to fast transducer adaptation in auditory hair cells,'' J. Neurosci. **20**, 7131–7142.

Sachs, F., and Lecar, H. (**1991**). ''Stochastic models for mechanical transduction,'' Biophys. J. **59**, 1143–1145.

Salje, E. (**1990**). *Phase Transitions in Ferroelastic and Co-elastic Crystals* (Cambridge U. P., Cambridge, England).

Sukharev, S. I., Sigurdson, W. J., Kung, C., and Sachs, F. (**1999**). ''Energetic and spatial parameters for gating of the bacterial large conductance mechanosensitive channel, m_{scL},'' J. Gen. Physiol. **113**, 525–540.

Responses to cochlear normalized speech stimuli in the auditory nerve of cat

Alberto Recio[a] and William S. Rhode
Department of Physiology, University of Wisconsin, Madison, Wisconsin 53706

Michael Kiefte
School of Human Communication Disorders, Dalhouise University, Halifax, Nova Scotia B3H 1R2, Canada

Keith R. Kluender
Department of Physiology and Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706

(Received 14 December 2001; revised 18 February 2002; accepted 22 February 2002)

Previous studies of auditory-nerve fiber (ANF) representation of vowels in cats and rodents (chinchillas and guinea pigs) have shown that, at amplitudes typical for conversational speech (60–70 dB), neuronal firing rate as a function of characteristic frequency alone provides a poor representation of spectral prominences (e.g., formants) of speech sounds. However, ANF rate representations may not be as inadequate as they appear. Here, it is investigated whether some of this apparent inadequacy owes to the mismatch between animal and human cochlear characteristics. For all animal models tested in earlier studies, the basilar membrane is shorter and encompasses a broader range of frequencies than that of humans. In this study, a customized speech synthesizer was used to create a rendition of the vowel [ɛ] with formant spacing and bandwidths that fit the cat cochlea in proportion to the human cochlea. In these vowels, the spectral envelope is matched to cochlear distance rather than to frequency. Recordings of responses to this cochlear normalized [ɛ] in auditory-nerve fibers of cats demonstrate that rate-based encoding of vowel sounds is capable of distinguishing spectral prominences even at 70–80-dB SPL. When cochlear dimensions are taken into account, rate encoding in ANF appears more informative than was previously believed. [DOI: 10.1121/1.1468878]

PACS numbers: 43.64.Pg, 43.64.Sj [LHC]

I. INTRODUCTION

Numerous studies have explored encoding of speech sounds in the peripheral and central auditory system via *in vivo* studies in animal models (e.g., Sachs and Young, 1979; Blackburn and Sachs, 1990; Recio and Rhode, 2000). Investigations of auditory-nerve fiber (ANF) responses reveal that a representation of vowel spectra based on the individual firing rates of ANFs is inadequate for transmitting information known to be conveyed in humans at levels typical of conversational speech (60–70 dB). At higher presentation levels, firing rates can saturate, resulting in nearly flat frequency-rate curves that do not resolve frequencies in complex sounds. In part, this is due to the limited dynamic range of mid- and high-spontaneous fibers (typically less than 24 dB; Schalk and Sachs, 1979). Saturation alone does not explain all of the apparent lack of frequency selectivity because peak response rates often are substantially lower than the saturation level in response to characteristic-frequency (CF) tones. It has been suggested that two-tone suppression, by which energy in one region of the spectrum (e.g., first formant F_1) suppresses responses to energy in another region (e.g., second and third formants, F_2 and F_3), could account for the lower response rates (Sachs and Young, 1979). Inter-fiber variability, which can be due to differences in threshold or dynamic range, can account for some of the poor performance of the rate code at high stimulus levels (Miller *et al.*, 1999). When much of this variability is factored out, as in the case of analyzing rate differences in responses to two stimuli, rate becomes more efficient for encoding differences between those two stimuli even at 70-dB SPL (Conley and Keilson, 1995).

Responses to vowel sounds similar to those in ANF have been recorded in neurons of cochlear nucleus (CN: Blackburn and Sachs, 1990; Recio and Rhode, 2000). Neuron types differed in their responses. Medium- and high-spontaneous nerve fibers provide a poorer rate-based representation of vowel spectra than low-spontaneous fibers, which can encode spectral peaks of speech sounds even at 70–80 dB SPL (Sachs and Young, 1979; Blackburn and Sachs, 1990). In CN, transient chopper neurons also maintain an accurate rate-based representation of spectra of vowels (Blackburn and Sachs, 1990; Recio and Rhode, 2000).

Whereas studies cited above included close attention to the fidelity of computer-generated vowels, they overlooked some potentially important differences between human auditory systems and those of the animals under investigation. Cat and chinchillas, which have been especially prominent animal subjects in earlier studies, have a considerably shorter cochlea than that of humans (e.g., 25 and 17.5 mm for cat and chinchilla, respectively, versus 35 mm for human). Furthermore, cats are sensitive to a considerably broader range

[a]Electronic mail: recio@physiology.wisc.edu

of frequencies than humans (up to 60 kHz compared to 20 kHz for humans). These characteristics are significant when one wishes to understand neural representation of speech sounds. For example, positions of the center frequencies of the first three spectral prominences (F_1, F_2, and F_3) of the vowel [ε] (as in "bet") have been synthesized at 512, 1792, and 2432 Hz, respectively, corresponding closely to means observed from natural productions (Peterson and Barney, 1952; Sachs and Young, 1979). According to Greenwood's (1990) estimates, in a human cochlea the distances to the apex for each of those frequencies are 10.21, 17.88, and 19.93 mm, respectively. The same center frequencies in the cat will be located about 3.38, 8.03, and 9.38 mm from the apex. Consequently, the distance between the first two formant peaks (F_1, F_2) in the cat and chinchilla, 4.65 and 4.14 mm, respectively, is smaller than that measured in humans (7.67 mm). Although one should think of the above distances as approximations—particularly in the case of humans and chinchillas, for which accurate cochlear maps are not available—basilar-membrane (BM) length in both cats and chinchillas is unquestionably shorter relative to humans.

In contrast to the frequency range of cat, the effective frequency range of human sensitivity is considerably smaller. Particularly for the range of frequencies used in speech, human hearing benefits from a longer cochlea that encodes a narrower range of frequencies. We reason that smaller cochlear distances in which vowel sounds are mapped into the cat and chinchilla cochleae may, in part, be responsible for the lack of good spectral resolution observed in rate-based representations of vowels by ANFs.

To test this hypothesis, a modified speech synthesizer (Kiefte *et al.*, 2002) was used to create a cochlear normalized version of the vowel [ε] based on the dimensions of the cat basilar membrane. This synthesis separated spectral peaks (formants) with respect to absolute cochlear distance rather than frequency. In addition, formant bandwidth was scaled to span equal numbers of inner-hair cells for cat and human. The vowel [ε] was selected because its neuronal representation has been the object of many studies (e.g., Sachs and Young, 1979; Young and Sachs, 1979; Blackburn and Sachs, 1990; May *et al.*, 1998; Recio and Rhode, 2000).

II. METHODS

A. Animal preparation

Data were obtained from 11 adult cats weighing between 2.5 and 6 kg. Cats were initially anesthetized with an intraperitoneal injection of pentobarbital (50 mg/kg). Additional smaller doses were administered through a catheter inserted in the femoral vein. A thermostatically controlled heating blanket maintained body temperature at 37 °C. After insertion of a tracheal cannula, the left ear was removed and the bulla was vented with 20 cm of 1-mm plastic tube. Removal of the overlying cerebellum made visualization of the cochlear nucleus possible and gentle retraction of the CN, using saline-soaked cotton balls, then exposed the auditory nerve. After covering the nerve with warm agar, a chamber was mounted over the skull opening and filled with mineral oil. The chamber was then sealed with a glass disk holding the electrode. KCl-filled micropipettes, with impedances of 30–70 MΩ, were used to record single-unit activity.

The protocol for these experiments was approved by the Animal Care Committee of the University of Wisconsin—Madison and meets NIH guidelines.

B. Acoustic stimuli, experimental protocol, and data analysis

A modified Radio Shack Super Tweeter (Chan *et al.*, 1993) served as the transducer for these stimuli. After calibration of the acoustic system (100–30 000 Hz, in 100-Hz steps) using a Bruel and Kjaer 0.5-in. condenser microphone, stimuli were digitally compensated for the transfer function of the acoustic system before being presented.

Vowel stimuli, original (human) and cochlear-normalized for cat, were synthesized using the program described in Kiefte *et al.* (2002). For human vowel $[\varepsilon]_H$, fundamental frequency (f_0) was 128 Hz, and center frequencies of F_1, F_2, and F_3 were set at 512, 1792, and 2432 Hz, respectively. Another version of the vowel was created, the cat $[\varepsilon]_C$, for which cochlear distance between formant peaks was held constant to the human counterpart. In creating this frequency mapping, values of f_0 (128 Hz) remained the same in both versions of the vowel. Formant frequencies for the cochlear-normalized version were obtained by first determining the cochlear position in millimeters of each formant based on Greenwood's (1990) equation

$$x = (1/a) * \log_{10}(F/A + k), \tag{1}$$

where a, A, and k are parameters estimated from human cochlear frequency position data, x is the distance along the basilar membrane from the apex, and F is the best frequency at location x. The inverse of this function using parameters appropriate for the cat cochlea is then used to estimate the normalized frequencies. In addition, normalized F_1 was adjusted downwards to 512 Hz by translating the vowel spectrum as a function of cochlear distance by an appropriate constant term in millimeters (Kiefte *et al.*, 2002). This ensured that the peak of F_3 would not be too close to the base of the basilar partition. As mentioned above, the cochlear distance between F_1 and F_2 stimulus in humans is 7.68 mm. In the cat's cochlea, keeping F_1 at 512 Hz, the value of F_2 needed to preserve the same distance is approximately 3508 Hz. Using the same procedure, F_3 would be approximately 5391 Hz. However, these values were rounded to the nearest harmonic frequency—i.e., 3456 and 5376 Hz for F_2 and F_3, respectively. Figure 1 shows the amplitude Fourier transform of the human (top panel) and cat (lower panel) vowels.

Duration of cat $[\varepsilon]_C$ and human $[\varepsilon]_H$ vowels was 100 ms. Stimuli were presented 100 times at a rate of 2 per s (400-ms intervening silence) at three levels: 30-, 50-, and 70-dB SPL. In the last set of experiments, stimuli were presented only at 80-dB SPL. All the stimuli were windowed using a raised cosine envelope of 5 ms at each end of the waveform. Cat and human vowels were played at a rate of 100 000 samples per second, using a 16-bit D/A.

During data acquisition, a response-area plot was initially obtained to determine best frequency (BF), threshold, and spontaneous rate (SR) of the fiber. Poststimulus time

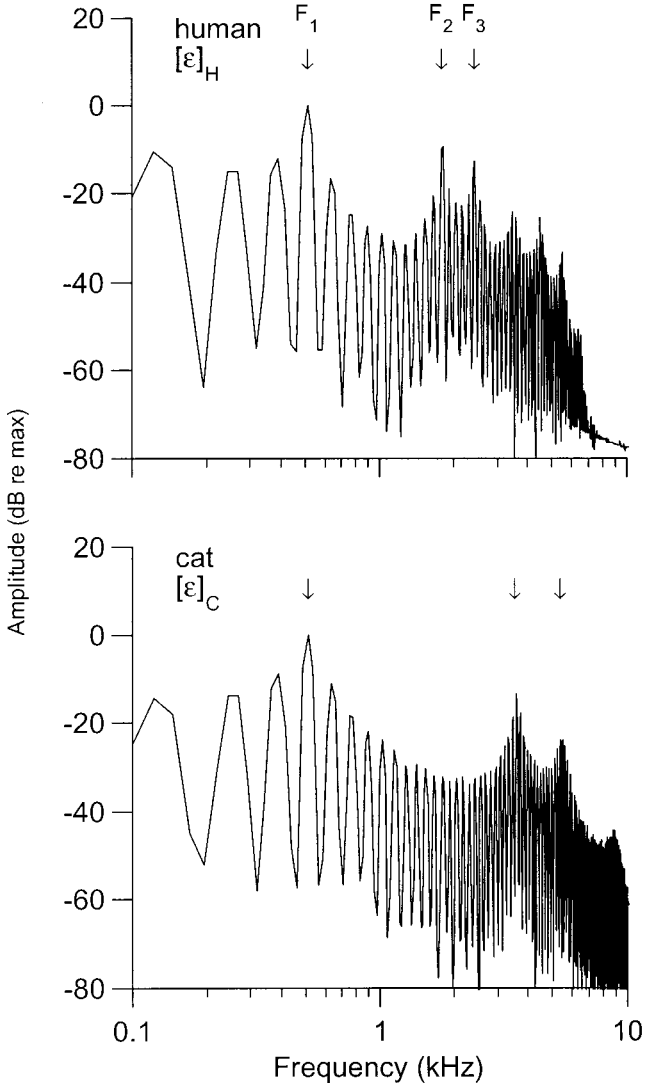


FIG. 1. Fourier transform amplitudes of human $[\varepsilon]_H$ (top panel) and cat $[\varepsilon]_C$ (low panel) sounds. Original waveforms were multiplied by a Hamming window ($N=4096$) and then Fourier transformed. Localizations of F_1–F_3 are indicated by arrows. Sampling rate=100 000 samples/s.

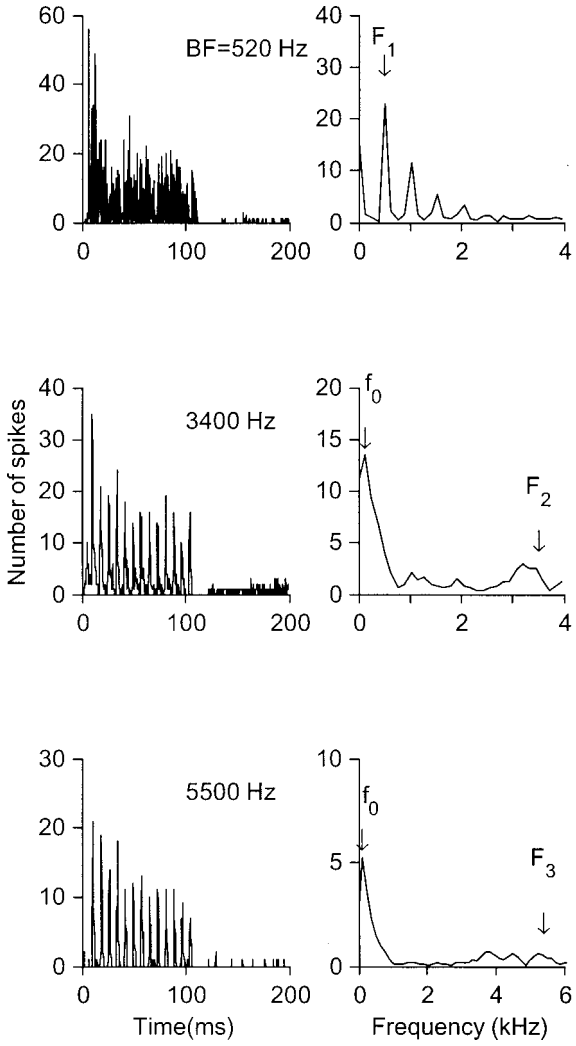


FIG. 2. Poststimulus time histograms (left column) with their corresponding Fourier transform amplitudes (right column) of the responses of three auditory-nerve fibers (ANFs) to the cat vowel $[\varepsilon]_C$. Fourier transform amplitudes were derived from cycle histograms, obtained with a period =128 Hz (f_0), computed from the neural responses. Unit's BF is indicated in each row. Stimuli (duration=100 ms, level=50 dB SPL) were presented 100 times with a repetition period of 0.5 s.

histograms (PSTHs) were routinely obtained using a 50-ms tone with frequency at BF. Rate-intensity curves using a 100-ms BF tone were also obtained to determine the maximum firing rate (MFR). Spikes were stored in the computer for later processing, which consisted mostly of determining the firing rate evoked by the speech sound. Firing rates evoked by the stimuli were estimated during the last 80 ms of the stimulus duration. To facilitate comparisons with previous work (e.g., Sachs and Young, 1979), the driven rate (rate evoked by speech sound-SR) and normalized rate (driven rate/[MFR−SR]) were also obtained for every neuron. For most neurons, the range of values of the normalized rate falls between 0 and 1. Whereas a normalized rate of 1 indicates that the neuron is driven by the speech sound at a rate identical to the maximum firing rate, a normalized rate of 0 indicates a lack of response by the neuron. Negative normalized rates can be a consequence of firing rates below the spontaneous activity level (rate suppression). Most of the analyses and plotting of figures were done in MATLAB.

Based on level of spontaneous activity, units were classified as low (SR<1 spikes/s), medium (1≤SR<18 spikes/s), and high-spontaneous (SR≥18 spikes/s) fibers. Data reported here are from 282 units with BFs<10 kHZ.

III. RESULTS

A. Responses of individual fibers

Because ANF responses to the human vowel $[\varepsilon]_H$ have been described elsewhere (e.g., Sachs and Young, 1979; Recio and Rhode, 2000), we will concentrate on responses to the vowel normalized to the cat cochlea. Figure 2 shows the responses, in the form of PSTHs and Fourier transform amplitudes, of three ANFs to cat $[\varepsilon]_C$ at 50-dB SPL. BFs of the fibers in the top, center, and bottom rows are near-center frequencies of F_1, F_2, and F_3, respectively. The response of low-BF neurons (i.e., neurons with BF around F_1) shows synchronization to the peak harmonic of F_1 (512 Hz), with additional phase locking to successive harmonics of 512 Hz

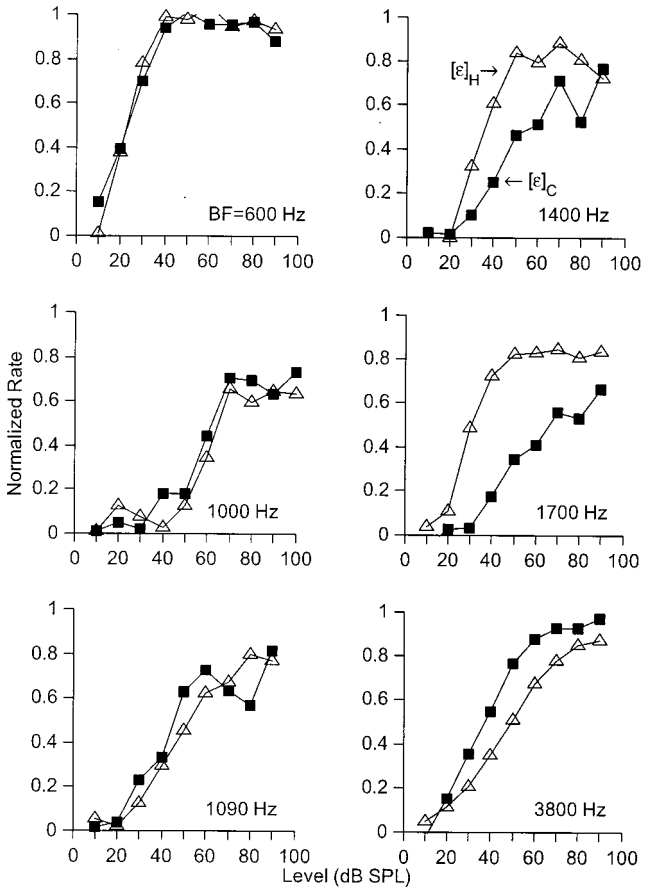


FIG. 3. Normalized rates versus stimulus level curves obtained from responses of six ANFs. Open triangles indicate responses to the human $[\varepsilon]_H$ sound and filled squares show the responses to $[\varepsilon]_C$. Responses to vowels and BF tones were obtained from ten stimulus presentations (duration=100 ms) with a repetition rate of 0.5 s. Normalized rates were computed using steady-state (20–100-ms) spike count.

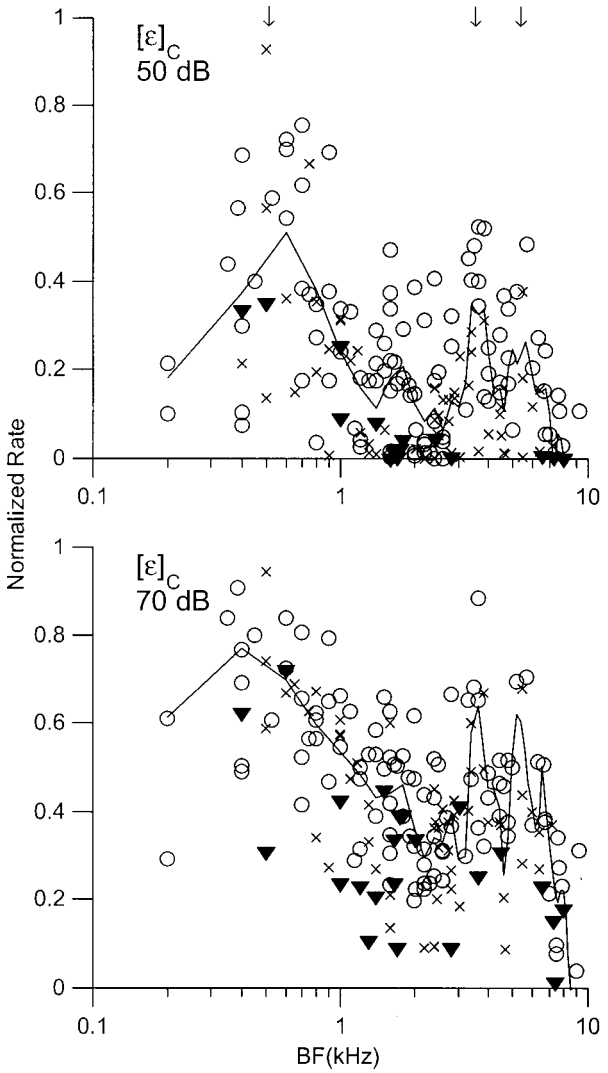


FIG. 4. Normalized rates obtained from steady-state responses of a population of 181 ANFs to $[\varepsilon]_C$ at two stimulus levels (50- and 70-dB SPL). Open circles, crosses, and filled inverted triangles indicate the responses of high-, medium-, and low-spontaneous ANFs, respectively. Continuous lines show the average normalized rate at a given BF, using frequency bins=250 Hz. Arrows indicate the value of F_1-F_3.

(i.e., 1024, 1536, and 2048 Hz). PSTHs of high-BF neurons (i.e., with BFs around F_2 or F_3) include peaks separated from each other at an interval equal to the period corresponding to f_0. This is confirmed in the frequency domain where the largest spectral peak occurs at f_0. Phase locking to the frequencies of F_2 (3456 Hz) is almost nonexistent.

Normalized rate curves versus stimulus level in response to $[\varepsilon]_H$ and $[\varepsilon]_C$ were also obtained, as shown in Fig. 3. For neurons with BF near F_1, such as the one in the top-left panel, normalized rates in response to human and cat vowels are nearly identical and reach a maximum value of 1. This is expected because the center frequency of F_1 is set to the same value for both $[\varepsilon]_H$ and $[\varepsilon]_C$. Even at BFs around 1 kHz (center and bottom left panels), normalized rates in response to $[\varepsilon]_H$ remain similar to those obtained in responses to $[\varepsilon]_C$. Maximum values of normalized rates, however, are around 0.8. Whether the lower maxima are due to the stimulus not being of sufficiently high amplitude or to rate suppression is unclear. Suppression by F_2, if any, appears to be independent of the center frequency (or cochlear position) of that formant. Normalized rates for neurons with BF=1400 and 1700 (top and center right) are usually higher when driven by the $[\varepsilon]_H$ (open triangles) than by $[\varepsilon]_C$, presumably as a consequence of proximity of these BFs to F_2 of $[\varepsilon]_H$ (1792 versus 3456 Hz for $[\varepsilon]_C$). Dynamic range (~60 dB) of the responses for some of those neurons to $[\varepsilon]_C$ (filled squared) is also larger than the dynamic range obtained from responses to BF tones. The neuron with BF=3800 (lower right) is driven at a higher rate by $[\varepsilon]_C$ than $[\varepsilon]_H$. This is not surprising because the BF of that fiber is near F_2 of $[\varepsilon]_C$ (3456 Hz).

B. Population analysis

Figure 4 shows the responses of a population of ANFs to $[\varepsilon]_C$ at two levels. Responses of high-spontaneous fibers (open circles) yield higher normalized rates than medium- and low-spontaneous units. This was expected because high-spontaneous units have lower thresholds than medium- and low-spontaneous fibers (Liberman, 1978). Continuous lines show the average discharge rate for medium- and high-spontaneous fibers only. Individual points constituting the lines represent the average normalized rate among those neu-

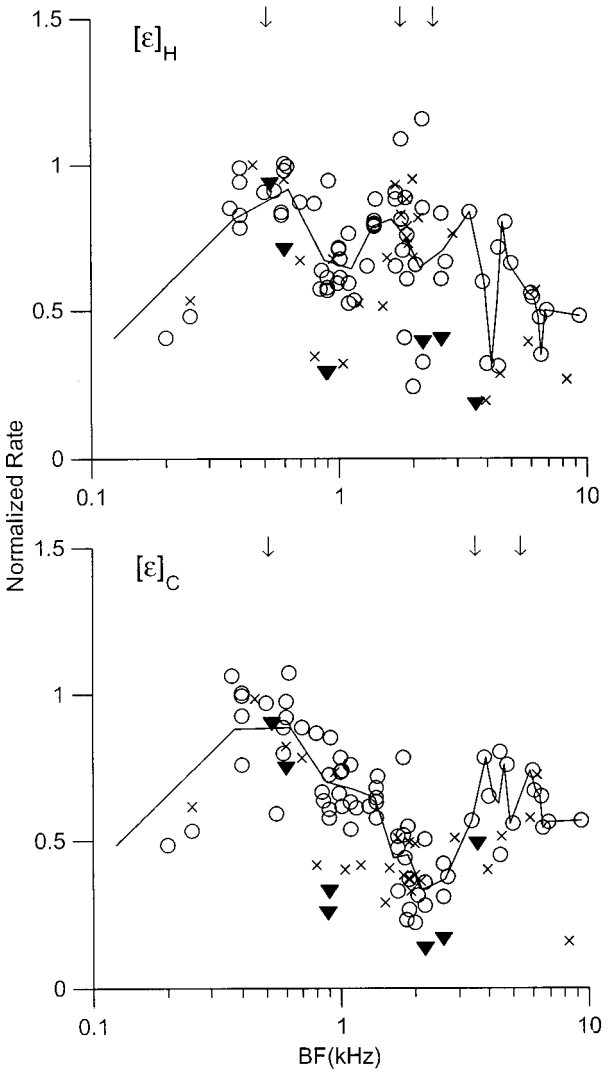


FIG. 5. Normalized rates obtained from steady-state responses of 100 ANFs to $[\varepsilon]_H$ (top panel) and $[\varepsilon]_C$ (lower panel) at 80 dB DPL. Symbols, arrows, and continuous lines represent the same information as in Fig. 4.

rons whose BF is within 125 Hz of the abscissa value (i.e., a 250-Hz window). This was done to facilitate comparisons with the work of Sachs and Young. At both stimulus levels, average rates indicate the presence of three distinct peaks, one for each of the first three formants of the vowels. Sachs and Young (1979) show similar results for responses to a 48-dB vowel sound. In contrast to the results found in the present study, Sachs and Young were unable to show distinct spectral peaks at higher stimulus levels.

Responses to 80-dB vowel sounds obtained from a different neuronal population are shown in Fig. 5, with continuous lines indicating average values. Continuous lines were obtained as described above. Normalized rates obtained from responses to $[\varepsilon]_H$ provide a poor representation of the spectrum of the vowel (i.e., there is little spectral definition), a finding that agrees with the results of Sachs and Young (1979). When the same group of neurons was stimulated using the cochlear-normalized vowel, however, normalized rates show the presence of two spectral peaks, one at the location of F_1 and the other in the region of F_2–F_3, with a distinct trough in between. The variability in thresholds of units shown in Fig. 5 was smaller than the variability of the neurons in Fig. 4, which may account for the larger fluctuations observed in normalized rates of Fig. 4.

IV. DISCUSSION

The results of this study show that, when the spectral composition of a vowel sound is adjusted to map onto the animal cochlea as a function of cochlear distance as it would fit the human cochlea, rate-based encoding of speech stimuli provides more information than has previously been assumed, even at 70–80-dB SPL. Relatively better preservation of spectral definition is maintained in spite of potential suppressive effects at the level of the basilar membrane. In fact, rate suppression in fibers with BFs near the spectral trough between F_1 and F_2 appears to be independent of the center frequency of F_2. This can be observed from results from Fig. 5, in which normalized rates values around 1–1.5 kHz are similar in responses to human and cat vowel sounds.

While normalizing spectral composition of speech sounds to nonhuman cochlear dimensions has given new insights into potential encoding of speech in the human auditory system, this spectral adjustment introduces at least one disadvantage. Upward displacement of spectral peaks moves formant center frequencies beyond the range of phase locking. For example, F_2 for $[\varepsilon]_C$ is at 3456 Hz and F_3 is at 5376 Hz. Spectral analysis of responses of neurons with BFs in these regions show a very small spectral peak at those frequencies. For animals such as chinchilla, keeping F_1 at 512 Hz results in a shift in the frequency to F_2 to approximately 4.7 kHz, where phase locking is nearly zero. Although ANFs can phase lock to signals with frequencies in the 2–4-kHz range, the significance of this phase locking is not yet clear. Whereas there is an enhancement of synchrony in the CN to signals below 1 kHz (Joris *et al.*, 1994), there is also deterioration of synchronization indices for CN responses to frequencies above 1–2 kHz relative to those measured in ANFs (Rhode and Smith, 1986). Further decrease in synchrony occurs at higher levels in the auditory system, such as in the inferior colliculus and auditory cortex.

A final potential disadvantage is the implicit assumption that basilar-membrane mechanics is the same in the apex as in the base of the cochlea, or in intermediate points. With the exception of early measurements in the squirrel monkey (Rhode, 1971), direct measurements of BM motion exist only in the base and apex of certain animals. Results from the apex (Rhode and Cooper, 1996) indicate that the amount of nonlinearity is smaller than in the base, yet frequency sensitivity is broader near the apex than at the base. Indirect measurements of BM motion, such as those obtained from *revcor* functions of ANFs (Carney *et al.*, 1999) also show differences in impulse responses of fibers near the apex (BF $<$500 Hz) relative to those measured from ANFs with BF$>$1 kHz. Therefore, it is conceivable that mechanics for the cochlear location with BF in the human at 1792 Hz might not be strictly comparable to the corresponding location in cat cochlea (3456 Hz).

It is important to note that these shortcomings may be problems for cat or chinchilla, but this does not imply a problem for humans. For cat and chinchilla, one cannot have

cochlear distances between formants equal to those for humans and have spectral peaks for F_2 and F_3 within the region of reliable phase locking. For humans, with shorter frequency range and longer cochleae, it is possible to maintain cochlear distance and have phase locking, too. Of course, little is known about the mechanics of living human cochleae.

Limitations aside, the present results suggest that rate encoding of human speech sounds may be significantly more robust in the human ANF than would be expected based upon earlier measures with human speech sounds presented to nonhuman auditory systems.

ACKNOWLEDGMENTS

Thanks to Laurel Carney and two anonymous reviewers for their comments. Work supported by NIH Grant Nos. NS-17590 and DC-04072.

Blackburn, C. C., and Sachs, M. B. (**1990**). ''The representation of the steady-state vowel sound /ε/ in the discharge patterns of cat anteroventral cochlear nucleus neurons,'' J. Neurophysiol. **63**, 1191–1212.

Carney, L. H., McDuffy, M. J., and Shekhter, I. (**1999**). ''Frequency glides in the impulse responses of auditory nerve fibers,'' J. Acoust. Soc. Am. **105**, 2384–2391.

Chan, J. C. K., Musicant, A. D., and Hind, J. E. (**1993**). ''An insert earphone system for delivery of spectrally shaped signals for physiological studies,'' J. Acoust. Soc. Am. **93**, 1496–1501.

Conley, R. A., and Keilson, S. E. (**1995**). ''Rate representation and discriminability of second formant frequencies for /ε/-like steady-state vowels in cat auditory nerve,'' J. Acoust. Soc. Am. **98**, 3223–3234.

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

Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. C. T. (**1994**). ''Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency,'' J. Neurosci. **71**, 1022–1036.

Kiefte, M., Kluender, K. R., and Rhode, W. S. (**2002**). ''Synthetic speech stimuli spectrally normalized for nonhuman cochlear dimensions,'' ARLO **3**, 41–46.

Liberman, M. C. (**1978**). ''Auditory-nerve responses from cats raised in a low-noise chamber,'' J. Acoust. Soc. Am. **63**, 442–455.

May, B. J., LePrell, G. S., and Sachs, M. B. (**1998**). ''Vowel representation in the ventral cochlear nucleus of the cat: Effect of level, background noise, and behavioral state,'' J. Neurophysiol. **79**, 1755–1767.

Miller, R. L., Calhoun, B. M., and Young, E. D. (**1999**). ''Discriminability of vowel representations in cat auditory-nerve fibers after acoustics trauma,'' J. Acoust. Soc. Am. **105**, 311–325.

Peterson, G., and Barney, H. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.

Recio, A., and Rhode, W. S. (**2000**). ''Representation of vowel stimuli in the ventral cochlear nucleus of the chinchilla,'' Hear. Res. **146**, 167–184.

Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.

Rhode, W. S., and Cooper, N. P. (**1996**). ''Nonlinear mechanics in the apical turn of the chinchilla cochlea *in vivo*,'' Aud. Neurosci. **3**, 101–121.

Rhode, W. S., and Smith, P. H. (**1986**). ''Encoding timing and intensity in the ventral cochlear nucleus of the cat,'' J. Neurophysiol. **56**, 261–286.

Sachs, M. B., and Young, E. D. (**1979**). ''Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate,'' J. Acoust. Soc. Am. **66**, 470–479.

Schalk, T. B., and Sachs, M. B. (**1979**). ''Nonlinearities in auditory-nerve fiber responses to bandlimited noise,'' J. Acoust. Soc. Am. **67**, 903–913.

Young, E. D., and Sachs, M. B. (**1979**). ''Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers,'' J. Acoust. Soc. Am. **66**, 1381–1403.

Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited[a)]

Ewan A. Macpherson[b)] and John C. Middlebrooks
Kresge Hearing Research Institute, University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109-0506

(Received 24 September 2001; accepted for publication 26 February 2002)

The virtual auditory space technique was used to quantify the relative strengths of interaural time difference (ITD), interaural level difference (ILD), and spectral cues in determining the perceived lateral angle of wideband, low-pass, and high-pass noise bursts. Listeners reported the apparent locations of virtual targets that were presented over headphones and filtered with listeners' own directional transfer functions. The stimuli were manipulated by delaying or attenuating the signal to one ear (by up to 600 μs or 20 dB) or by altering the spectral cues at one or both ears. Listener weighting of the manipulated cues was determined by examining the resulting localization response biases. In accordance with the Duplex Theory defined for pure-tones, listeners gave high weight to ITD and low weight to ILD for low-pass stimuli, and high weight to ILD for high-pass stimuli. Most (but not all) listeners gave low weight to ITD for high-pass stimuli. This weight could be increased by amplitude-modulating the stimuli or reduced by lengthening stimulus onsets. For wideband stimuli, the ITD weight was greater than or equal to that given to ILD. Manipulations of monaural spectral cues and the interaural level spectrum had little influence on lateral angle judgements. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1471898]

PACS numbers: 43.66.Qp [LRB]

I. INTRODUCTION

The locations of sound sources are not mapped directly at the sensory periphery. Instead, locations must be derived by combining acoustical cues that result from the interaction of incident sound waves with the external ears, head, and upper body. The acoustical cues for sound localization were explored as early as the end of the 19th century by the British physicist Lord Rayleigh (Strutt, 1907), among others. Rayleigh worked primarily with pure-tone stimuli produced by vibrating tuning forks. He determined that the primary cue to the lateral positions of sources with frequencies >500 Hz was the interaural difference in sound pressure levels (ILDs) resulting from acoustic shadowing by the head. At lower frequencies, however, the wavelength of sound is much larger than the diameter of the head, and ILDs are negligible. Through the ingenious use of a pair of mistuned low-frequency tuning forks, Rayleigh demonstrated compellingly that human listeners are sensitive to interaural differences in the ongoing phase of low-frequency sounds and, thus, that interaural time differences (ITDs) could provide cues to the lateral positions of low-frequency sources. Rayleigh's understanding of the localization of tones in the lateral dimension has come to be known as the ''Duplex Theory'' of sound localization and has been substantiated in numerous psychophysical and physiological studies. Rayleigh also appreciated that subjects could not discriminate the front-versus-back locations of pure-tone stimuli, but that such front/back discrimination was possible for ''sounds of other character'' (i.e., sounds with broader bandwidths). Research in the past few decades has filled in some understanding of the cues for front/back and vertical localization, revealing the importance of spectral-shape cues provided by the direction-dependent filtering of broadband sounds by the external ears.

The Rayleigh Duplex Theory is quite satisfactory to explain left/right localization of tonal stimuli. Nevertheless, in the real world most sounds have bandwidths of several octaves, and a listener rarely is exposed to a pure tone. Three sets of observations led us to revisit the Duplex Theory in the context of complex (i.e., broadband) sounds.

First, lateralization studies have shown than listeners are sensitive to ITDs in high-frequency complex sounds. Sensitivity to ongoing time differences in simple sounds such as pure tones is limited to low frequencies by the loss of phase-locking in the auditory nerve at high frequencies and likely also by a lower-frequency cutoff in the binaural system. Indeed, psychophysical studies show that sensitivity to ongoing ITD is limited to frequencies below ~1.3 kHz for pure tones (Zwislocki and Feldman, 1956). Nevertheless, the auditory system can extract timing information from the envelopes of higher-frequency sounds that contain multiple frequency components, and listeners can detect ITDs in high-frequency complex sounds presented through headphones (e.g., Henning, 1974; Leakey *et al.*, 1958; McFadden and Pasanen, 1976).

Second, Wightman and Kistler (1972, 1997b) have investigated localization cues using virtual auditory space (VAS) techniques, which permit more-or-less independent manipulation of ITD, ILD, and spectral-shape cues. They demonstrated that ITDs dominate listeners' judgments of the location of broadband sound sources that contain low-

[a)]Portions of this work were previously presented in poster form at the 24th Annual Midwinter Research Meeting of the Association for Research in Otolaryngology in February 2001.

[b)]Electronic mail: emacpher@umich.edu

frequency components (Wightman and Kistler, 1992). In some cases, the influence of ITDs persisted even when stimulus spectra were limited to high frequencies, although the effect of high-pass filtering varied widely among listeners. Also, imposition of an interaural imbalance (i.e., an ILD of 10–20 dB) had surprisingly little impact on lateral location judgments by some listeners (Wightman and Kistler, 1997b).

Third, spectral-shape cues, which provide essential cues to the vertical and front/back location of broadband sounds, also vary with source azimuth and might contribute to the judgment of lateral localization. For instance, some congenitally monaural listeners, whose only cues to sound-source location derive from direction-dependent filtering by one external ear, show reasonably accurate localization in the horizontal dimension (Slattery and Middlebrooks, 1994). Whether normal binaural listeners similarly use spectral cues in determining lateral angle is unknown.

These three sets of observations raise questions about the applicability of the Rayleigh duplex model to localization of naturally occurring broadband sounds. Might ITDs contribute to or even dominate localization of high-frequency complex sounds? How salient are spectral-shape cues compared to interaural difference (i.e., ITD and ILD) cues? Although ILDs have an obvious impact on lateralization of high-frequency tones, do ILDs have any influence on localization of broadband sounds?

We addressed these questions by measuring localization of broadband, low-pass, and high-pass sounds presented using VAS techniques. We quantified the degrees to which cues addressed by the Duplex Theory (such as low-frequency ITD and high-frequency ILD) and other cues (such as spectral cues and high-frequency, envelope-based ITD) contributed to the lateral component of listeners' localization judgments. Apart from some influence of envelope-based ITDs in high-pass signals, we found that the localization of complex sounds in the lateral dimension accorded nicely with a more-than-century-old theory based on the localization of vibrating tuning forks.

II. METHODS

A. Subjects

Thirteen paid listeners (five female and eight male, ages 18–35 years, including the first author) were recruited from the student body of the University of Michigan and the staff of the Kresge Hearing Research Institute. All had extensive experience in free-field localization experiments. Only one (the first author, S18) had previous experience in localizing virtual auditory stimuli presented over headphones, but the others received several hours of practice in the virtual free-field localization task described in Sec. II D prior to the beginning of the present study. These practice sessions involved the localization of virtual broadband noise-burst targets. No feedback was provided. All listeners had normal hearing as defined by standard audiometric testing. Ten of the listeners (not including S18) participated in Experiments I and II-A. Two of these ten plus S18 and two additional listeners participated in Experiment II-B. Three of the ten plus S18 participated in Experiment III, and three from Experiment II-B participated in Experiment IV.

B. Directional transfer function measurements

In order to compute the filters necessary to synthesize the virtual free-field stimuli, measurements were made of each listener's directional transfer functions (DTFs). The details of this procedure are given by Middlebrooks (1999a). Briefly, 512-point, 50-kHz Golay codes (Zhou *et al.*, 1992) were presented from a loudspeaker positioned 1.2 m from the listener's head at 400 locations approximately evenly distributed in space around the listener's head. The responses to these excitation signals were recorded simultaneously by two miniature electret microphones (Knowles, model 1934) inserted approximately 5 mm into the listeners' ear canals. Head-related transfer functions (HRTFs) were extracted by cross-correlation of excitation and response, transformation to the frequency domain, and division by the previously measured loudspeaker transfer function. For each ear, DTFs were computed from the HRTFs by dividing by a complex common component computed for the set of HRTFs for that ear (Middlebrooks, 1999a). This nondirectional component was a combination of the ear canal resonance and the diffuse-field average response at the ear canal entrance. The DTFs were transformed into the time domain, yielding a set of directional impulse responses (DIRs) for each listener. The DIRs were used in the synthesis of the virtual free-field targets as described below.

C. Stimulus synthesis

Stimulus noise waveforms were computed on an Intel-based desktop personal computer using custom MATLAB scripts (The Mathworks). An inverse-Fourier-transform method was used to produce flat-spectrum, random-phase noise waveforms with the desired passband and duration sampled at 50 kHz. Raised-cosine (i.e., $\cos^2$) ramps of 1, 20, or 50 ms duration (depending on the stimulus condition) were applied to the onsets and offsets. The resulting waveform was convolved with the right- and left-ear DIRs corresponding to the desired target location. The stimuli were presented over ''diffuse-field equalized'' circumaural headphones (Sennheiser HD 265) at a sampling rate of 50 kHz using digital-to-analog converters, attenuators, and headphone amplifiers from Tucker-Davis Technologies (models DD1, PA4, and HB6, respectively). The stimulus level on each trial was equivalent to that of a free-field source at a sound pressure level of approximately 65 dB.

We did not attempt a rigorous equalization of the headphone response. Rather, the headphone response itself restored an approximation of the diffuse-field component removed in the computation of the DIRs, and the listener's own ear canal restored the ear canal resonance. We have discussed this approach previously and have shown that listeners can localize accurately in the VAS generated by this method (Middlebrooks, 1999b).

In Experiments I and II, the interaural time difference (ITD) and interaural level difference (ILD) cues naturally

TABLE I. Summary of cue manipulations for Experiments I–IV. The cues referred to are interaural time difference or interaural phase spectrum (ITD/IPS), interaural level difference or interaural level spectrum (ILD/ILS), and the directional transfer functions for the ears nearer to and farther from the source (DTF_{near} and DTF_{far}). Symbols indicate whether the cue corresponded to the original target location (□), to a biased location (■), or (in the case of the artificial DTFs used in Experiments III and IV) to no actual location (⋆).

Cue	Experiments I and II		Experiment III ILS bias	Experiment IV DTF_{near}, bias
ITD/IPS	■	□	□	□
ILD/ILS	□	■	■	□
DTF_{near}	□	□	□	■
DTF_{far}	□	□	⋆	⋆

present in the stimulus were manipulated by imposing a whole-waveform delay or attenuation on the signal at one of the ears. We refer to this procedure as *biasing* the stimulus and to the amount of ITD or ILD offset as the *imposed bias*. In Experiments III and IV, the stimuli were manipulated by modifying the DTF spectrum at one or both ears in order to bias the interaural level spectrum (ILS) or the DTF of the ear nearest the source. The cue manipulations for Experiments I–IV are summarized in Table I.

D. Localization procedure

The localization procedure was similar to that described by Middlebrooks (1999b). Listeners stood in the center of a darkened anechoic chamber, and at the beginning of each trial oriented towards a light-emitting diode (the centering LED) positioned at eye level 2 m directly in front of the listener. A trial was initiated by pressing a hand-held button. The centering LED was extinguished, and the listener's initial head position was measured by a head-mounted electromagnetic tracking device (Polhemus FASTRAK). Following a delay of 500 ms, the stimulus was presented over headphones. After hearing the virtual free-field stimulus, the listener oriented towards its perceived location, at which time a second button press triggered a second measurement of head orientation. If the initial head orientation deviated by more than 10 degrees in either azimuth or elevation from the centering LED, the trial data were discarded. Depending upon the condition, runs consisted of 79, 99, or 111 trials, and the longest were typically completed in 10–12 min. Listeners rested after every two to three runs, and typically completed six to eight runs in one session.

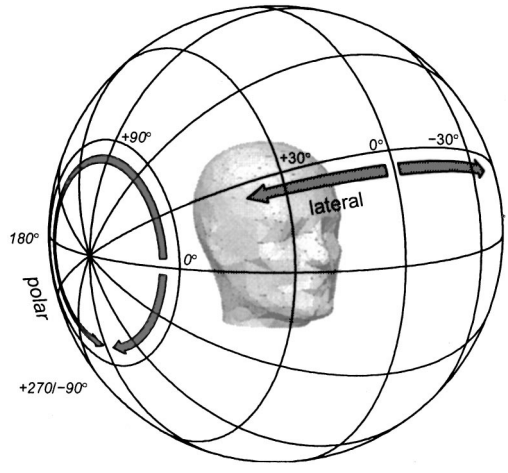


FIG. 1. Horizontal polar coordinate system. Lateral angle is the angle between the location and the median sagittal plane; positive values are to the listener's right. Polar angle combines elevation and front/back position. −90 degrees: below; 0 degrees: front; +90 degrees: overhead; +180 degrees: rear.

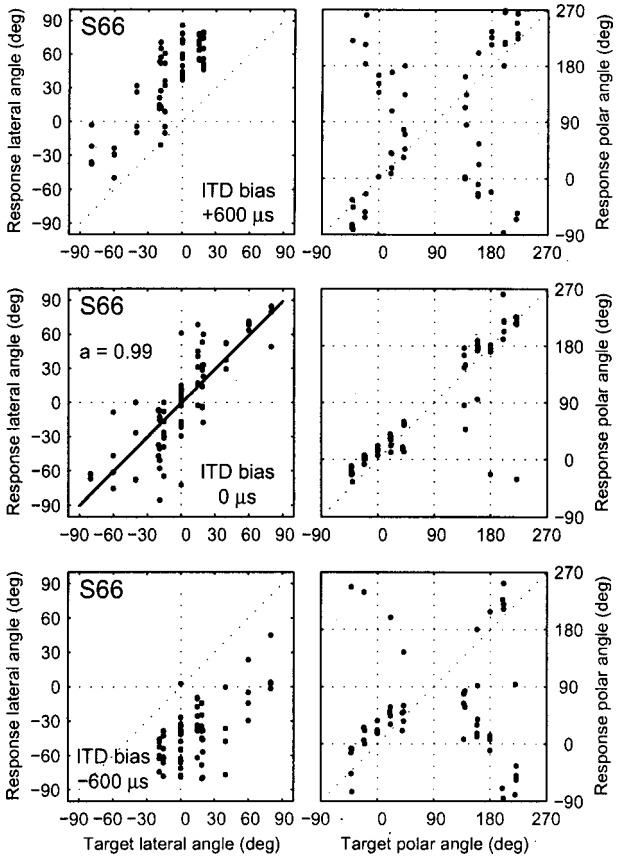


FIG. 2. Sample lateral- (left column) and polar-angle (right column) response data for ITD-biased wideband targets for listener S66. ITD bias: +600 μs (top), 0 μs (middle), −600 μs (bottom). Unbiased lateral angle gain, a, was computed as the slope of a linear fit to the unbiased stimulus lateral-angle responses (solid line). In the polar angle plots, accurate responses fell near the positive diagonal, and front/back reversed responses fell near the negative diagonal.

E. Data analysis

1. Lateral-polar coordinate system

Target and response location data were converted from the vertical polar (azimuth and elevation) coordinate system to the horizontal polar (lateral- and polar-angle) coordinate system (Fig. 1). Only the lateral angle data were analyzed in detail. Sample lateral and polar angle response data for wideband, ITD-biased stimuli are shown in Fig. 2. In this example, the ITD was biased by +600 μs (top row), 0 μs (middle row), or −600 μs (bottom row). Note that for these stimuli, positive (right-leading) and negative (left-leading) ITD biases shifted the listener's responses consistently towards the right and left sides, respectively. For this listener, applying ITD bias increased the rate of front/back confusions (more responses fell in the upper-left or lower-right quad-

rants), but the elevation component of the listener's polar-angle judgments remained reasonably accurate (i.e., responses fell near the positive or negative diagonals). The effect of bias on polar-angle responses is discussed further in Sec. III B 2.

2. Unbiased lateral angle gain

In order to reveal any detrimental effect on lateral angle localization caused by the restriction of stimulus bandwidth, an *unbiased lateral angle gain* (Hofman and Van Opstal, 1998; Macpherson and Middlebrooks, 2000) was computed for responses to the filtered, but unmanipulated (unbiased), targets in each stimulus set. The unbiased lateral angle gain was simply the slope of a linear fit to the target and response lateral angle data (Fig. 2).

3. ITD and ILD bias weights

We wished to assess the weighting or salience of the manipulated cue (ITD or ILD) in a manner which would permit meaningful comparisons between weights. Neither the classical time-intensity trading ratio (in units of μs/dB) nor ratios between angular displacements of judgments and magnitudes of imposed cue bias (yielding values in degrees/μs and degrees/dB) are useful in isolation in describing the relative effectiveness of ITD and ILD in localization. To use such values to make comparisons between the weighting of different cues, reference must also be made both to the naturally occurring relations between the physical cues and to the spatial disposition of the cues.

Our strategy was to derive *dimensionless* weights relating bias in responses to imposed bias in the underlying cue by (a) measuring the correspondence between the physical cues and lateral angle, and (b) using this relation to convert any shift in the angular response to a quantity expressed in the units of the manipulated cue itself.

As an example, consider a target location for which the natural ILD is 5 dB. If an ILD bias of 10 dB is imposed, the ILD in the presented stimulus will be 15 dB. If the listener responds at a location for which the natural ILD were also 5 dB (even if that were at a different lateral angle), we would conclude that the imposed bias had no effect, the observed bias would be 0 dB, and we would estimate the perceptual weighting of ILD to be close to 0. Conversely, if the listener responds at a location for which the natural ILD were 15 dB, we would conclude that the imposed bias was fully effective in shifting the perceived lateral angle, the observed bias would be 10 dB (equal to the imposed bias), and we would estimate the ILD bias weight to be close to 1. Responses at locations with intermediate ILDs would yield intermediate weights.

This conversion of bias in response location to bias in an underlying physical cue and the resulting derivation of a dimensionless weight permits the computation and comparison of perceptual weights on the ITD and ILD cues without concern about the physical correspondence of, or the auditory system's differential sensitivity to, the ITD and ILD cues. In addition, by incorporating the relation between cue and lateral angle into the procedure, the exact (and possibly nonmonotonic) relation between the interaural cues and lateral angle was rendered unimportant.

The following procedure was used to compute the bias weights. *First*, the value of the physical interaural cue at each of the 400 measured DTF locations was computed for each listener using procedures similar to those of Gaik (1993). For ILD, the energies in the right and left ear directional transfer functions were integrated over the stimulus passband (*wideband*, 0.5–16 kHz; *low-pass*, 0.5–2 kHz; *high-pass*, 4–16 kHz; see Secs. II C and III A) and their ratio represented in dB such that positive ILDs corresponded to higher intensity at the right ear.

To compute ITD, the DIR for each ear was passed through a gammatone filter bank (Slaney, 1994) with low-frequency channels at 600, 700, and 800 Hz, and high-frequency channels at 4, 4.5, and 5 kHz. These center frequencies were chosen because they produced the smoothest ITD-versus-azimuth functions across listeners. For the low-frequency channels, the ITD was taken from the lag of the peak in the cross-correlation of the left and right ear signals. For the high-frequency channels, the envelopes of the filter outputs were extracted using a Hilbert transform prior to cross-correlation in order to extract the group delay. This paralleled the loss of phase-locking and the onset of envelope following in the auditory nerve at high frequencies. If multiple peaks appeared in the cross-correlation, the one closest to the predicted ITD based on a spherical head model (Kuhn, 1987) was chosen. If no peak was found within 250 μs of the predicted value, the mean of the computed ITDs for neighboring locations was used. The median of the ITDs in the low-frequency channels was used as the ITD cue for low-pass and wideband stimuli, and the median of the high-frequency channel ITDs was used as cue for high-pass stimuli. Note that this process was intended as a means of measuring the physical ITD, not an attempt to model the extraction of ITD information by the auditory system.

Next, having associated a frequency-dependent ITD and ILD with each DTF location, an *observed cue bias* (ITD or ILD) was computed for each localization response in the manner illustrated in Fig. 3. The *natural cue*, c_{nat} (in μs or dB), was that present in the unmodified DTFs used in synthesizing the spatialized stimulus. The *observed cue*, c_{obs}, was that associated with the DTF location closest to the listener's response location. The observed cue bias was the difference between these values; bias$=c_{\mathrm{obs}}-c_{\mathrm{nat}}$. Trials in which the manipulated cue exceeded the range of the listener's measured ITD or ILD values were eliminated from the analysis. Overall, 10%–15% of trials were discarded for this reason.

Finally, the listener's weighting of the manipulated interaural cue was computed as the slope of the linear regression between the observed cue bias and the *imposed cue bias* (the magnitude of the added ITD, in μs, or ILD, in dB), as illustrated in Fig. 4. We refer to these dimensionless values as the $\mu s/\mu s$ and dB/dB weights. The standard error of the regression coefficient was taken as a measure of the uncertainty in the computed weight. If the manipulated cue had little influence on the listener's response, the response loca-

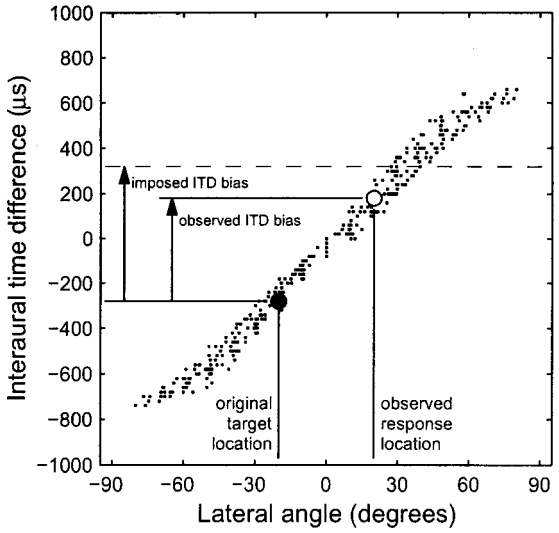


FIG. 3. Computation of observed cue bias. Small symbols show the measured low-frequency ITD for listener S92 as a function of lateral angle. Each small symbol represents a distinct DTF measurement location (unique lateral angle and polar angle combination). The large filled circle indicates the lateral angle and natural ITD of an example original target location. An ITD bias of +600 μs was applied. The large open circle indicates the lateral angle and natural ITD of the DTF measurement location closest to the listener's response (i.e., the observed ITD). The observed cue bias is the signed difference between the observed ITD and the natural target ITD.

tion was close to the original target location, the observed cue bias was close to zero in all trials, and the cue weight, W_{ITD} or W_{ILD}, was also close to zero (Fig. 4, upper-right and lower-left panels). Conversely, if the listener derived the judgment of lateral angle primarily from the manipulated cue, then the response was expected to lie at a location for which the natural cue was similar to the stimulus cue value. In such cases, the cue weight was close to 1 (Fig. 4, upper-left and lower-right panels).

III. EXPERIMENT I: WEIGHTING OF ITD AND ILD CUES IN LOW-PASS, HIGH-PASS, AND WIDEBAND NOISE

A. Stimuli and locations

In Experiment I, we measured listeners' weighting of the ITD and ILD cues to lateral angle under three passband conditions: *wideband*, 0.5–16 kHz; *low-pass* 0.5–2 kHz; and *high-pass*, 4–16 kHz. The target stimuli were 100-ms noise bursts with 1-ms raised-cosine onsets and offsets. Each stimulus set contained interleaved stimuli from four classes: (1) unfiltered (i.e., wideband) noise bursts, (2) filtered, unmanipulated (i.e., no imposed ITD or ILD bias) noise bursts, (3) filtered noise bursts with *medium* imposed cue bias (±300 μs ITD or ±10 dB ILD), and (4) filtered noise bursts with *large* imposed cue bias (±600 μs ITD or ±20 dB ILD). Some listeners also completed an additional set of ILD-bias conditions in which the medium and large biases of 10 and 20 dB were replaced by 4- and 8-dB biases, respectively. Analysis of these data showed that the computed weights were insensitive to which range of ILD biases was used, and the 4- and 8-dB ILD-bias data were not included in the following analysis.

Unfiltered targets, 36 in all, were placed at 10-degree increments in azimuth from −170 to +180 degrees with elevations of −30 or +30 degrees. Of the filtered, unbiased targets, 24 were placed at azimuths 0, ±20, ±160, and 180 degrees with elevations of ±20 and ±40 degrees (i.e., on or near the median plane, both in the front and rear, and above and below the horizontal plane). An additional 18 were

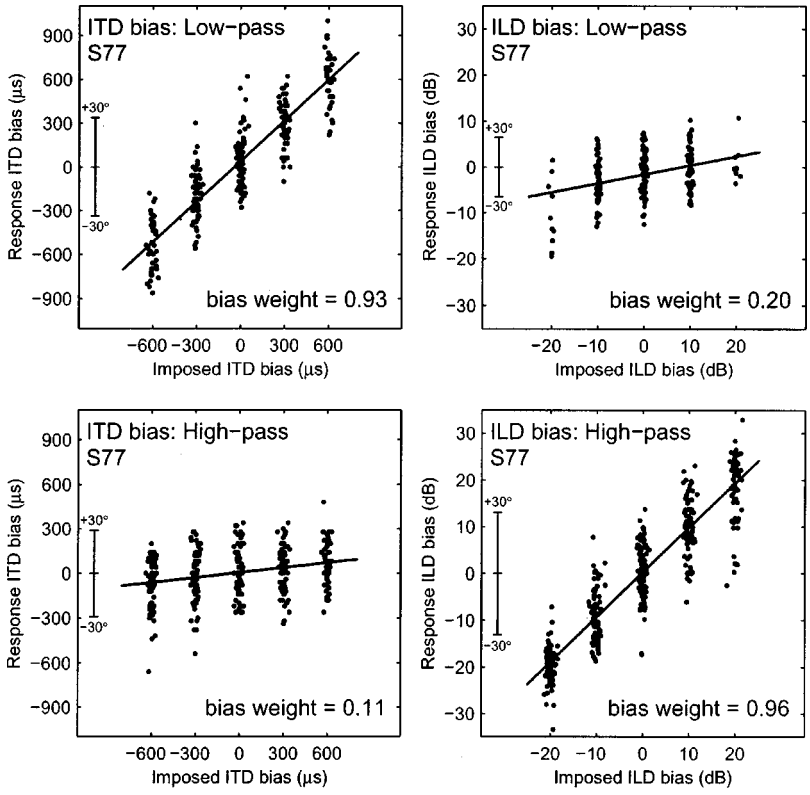


FIG. 4. Illustration of cue bias weight computation. Observed ITD and ILD cue biases are plotted against imposed cue bias for listener S77 in the low-pass (0.5–2 kHz) and high-pass (4–16 kHz) passband conditions. Upper-left: low-pass ITD-bias condition (bias=0, ±300, ±600 μs); upper-right: low-pass ILD bias condition (bias=0, ±10, ±20 dB); lower-left: high-pass ITD-bias condition; lower-right: high-pass ILD bias condition. The cue bias weight was the slope of a linear fit to these data. The ±30 degrees scale near the vertical axis indicates the natural variation of the manipulated cue within 30 degrees of the midline.

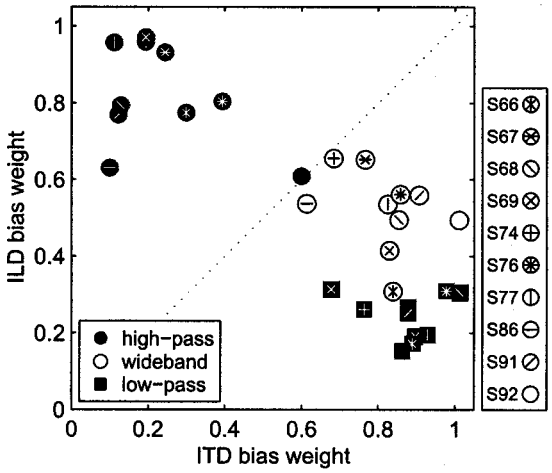


FIG. 5. Measured ITD- and ILD-bias weights for Experiment I. ILD-bias weight (vertical axis) is plotted against ITD-bias weight (horizontal axis) for each listener in the wideband, low-pass, and high-pass passband conditions.

placed in 20-degree increments of azimuth around the horizontal plane, for a total of 42 locations. To minimize the presentation of stimuli with ITD or ILD cues outside of the physiological range, the medium bias was not applied to targets more than 40 degrees away from the median plane on the side toward which the bias was applied, leaving 38 locations. Similarly, the large bias was not applied to targets more than 20 degrees from the median plane on the biased side, leaving 34 locations. In total there were six stimulus sets (3 passbands×2 cue-bias types) of 222 targets each. The targets in each set were presented twice in a shuffled order over the course of four blocks of 111 trials each. Blocks from different stimulus sets were intermixed within an experimental session but trials from different sets were not intermixed within blocks.

B. Results

1. Lateral angle responses

For each listener, the unbiased lateral angle gain was computed for the unmanipulated stimuli of each passband, and the cue weight was computed for each bias and passband condition. The lateral angle gain was unaffected by stimulus passband, and was close to unity for all listeners except for listener S69, who showed a mild decrement in lateral gain in the low-pass condition. Similar, near-unity, unbiased lateral angle gains were obtained for all other listeners in all conditions of Experiments I–IV. This result indicated that low-pass or high-pass filtering did not impair listeners' ability to judge accurately the lateral angle of the unbiased virtual free-field targets, and it increased our confidence that, in the biased conditions, the listeners also had access to usable cues to lateral angle in all passband conditions.

The cue bias weights are shown in the scatter plot of Fig. 5, in which ILD-bias weight is plotted against ITD-bias weight for each of the ten listeners in each passband condition.

The computed weights for the ITD and ILD cues were always greater than 0, and only in two cases exceeded 1 (for S68, $W_{\mathrm{ITD,LP}}=1.01$; for S92, $W_{\mathrm{ITD,WB}}=1.01$). Weights less than 1 were expected because the manipulated cue was placed in opposition to all the unmanipulated cues, which corresponded to a zero-bias target location. The unmanipulated cues thus opposed or diluted the effect of the biased cue. For clarity, the computed standard errors of the weights are not shown in Fig. 5, but they ranged from 0.01 to 0.04 for both ITD and ILD. The median standard error was ~0.025.

The relation between the ITD and ILD weights depended on the stimulus passband condition. In the wideband condition (open circles), ITD was weighted more heavily than ILD for all listeners (mean $W_{\mathrm{ITD,WB}}=0.82$, mean $W_{\mathrm{ILD,WB}}=0.52$), although the difference between the weights was small for listeners S67 and S86, and close to zero for S74. In the low-pass condition (filled squares), ITD weights were >0.8 for eight of the ten listeners, and were substantially higher than the ILD weights for all listeners (mean $W_{\mathrm{ITD,LP}}=0.88$, mean $W_{\mathrm{ILD,LP}}=0.24$). In the high-pass condition (filled circles), the relation between ITD and ILD weights was reversed from that observed in the low-pass case: ITD was weighted less heavily than ILD for all listeners except S92 (mean $W_{\mathrm{ITD,HP}}=0.24$, mean $W_{\mathrm{ILD,WB}}=0.82$). For all listeners, ITD weights were much lower than those observed in the wideband and low-pass conditions.

There were, however, marked individual differences in the weight given to the ITD cue in the high-pass passband condition. Six of the ten listeners (S68, S69, S74, S77, S86, S91) had small (<0.20) high-pass ITD weights, whereas the other four (S66, S67, S76, S92) had larger weights (0.24 to 0.60). Our high-pass stimuli were limited to frequencies above 4 kHz, and because the auditory system has virtually no representation of waveform fine structure at frequencies this high (Palmer and Russell, 1986), we presume that any ITD information utilized by these listeners must have been derived from the envelopes of the signals. We investigated the relative influence of the ITD cues provided by the onset and ongoing portions of the envelope in Experiment II.

2. Polar angle responses

We inspected the polar angle components of the listeners' responses under the conditions of restricted stimulus passband and imposed ITD or ILD bias. With no imposed bias, polar angle responses were similar to those observed in other free-field and virtual localization studies. In the wideband, unbiased condition, most listeners responded accurately in polar angle, but with rates of front/back confusion that varied among listeners. Errors of elevation and front/back position were most common for targets high above and behind the listener. The rate of front/back confusions increased for some listeners in the high-pass condition. Listeners' ability to localize the virtual stimuli in the vertical plane demonstrated that our stimulus synthesis was of sufficient quality to deliver accurate spectral cue information. We therefore presume that any influence of these spectral cues on apparent lateral angle (see Secs. V and VI) was similar to that occurring in real-world auditory environments. In the low-pass condition, no listeners were able to judge accurately the polar angle of the targets, and all made responses near the horizontal plane in either the front or rear hemi-

sphere for such filtered targets. This is a response pattern typically observed for low-pass stimuli (for example, Carlile *et al.*, 1999; Morimoto and Aokata, 1984).

The change in polar angle responses to ITD- or ILD-biased targets varied among listeners, but in general, the medium-magnitude biases (± 300 μs ITD or ± 10 dB ILD) produced little change from the unbiased condition. For approximately half of the listeners, large-magnitude biases (± 600 μs ITD or ± 20 dB ILD) substantially increased the rate of front/back confusions (as seen, for example, in Fig. 2) and caused mild compression of polar angle responses towards the horizontal plane. Wightman and Kistler (1992) have reported similar effects on vertical-plane localization for ITD-biased virtual stimuli. For the other listeners, even the large biases had little effect on polar responses. Sensitivity of polar response patterns to imposed ITD or ILD bias did not appear to be correlated with the weight given by individual listeners to the biased cue.

C. Discussion

1. Previous lateralization studies

Much of our knowledge about the auditory system's processing of interaural difference comes from so-called lateralization studies, in which stimuli are delivered over headphones without DTF filtering and their apparent positions lie inside the listener's head. In many respects, our derived weights for the ITD and ILD cues are consistent with the results of lateralization experiments employing both pure-tone and noise stimuli. We observed moderate to large weights on both ITD and ILD for wideband targets, although ITD was usually weighted more strongly; large ITD and small ILD weights for low-pass targets; and small ITD and large ILD weights for high-pass targets.

The intracranial position of diotic wideband noise can be displaced from the midline by the application of either ITD or ILD, and can be fully shifted to one side by either cue; similar sensitivities to ITD and ILD exist for low-frequency tones and low-pass noise (Blauert, 1997; Pinheiro and Tobin, 1969). ILD can displace the images of high-frequency tones (Fedderson *et al.*, 1957) and bands of noise (Simon and Aleksandrovsky, 1997), but high-frequency tones above ~ 1.3 kHz cannot be lateralized on the basis of ongoing ITD (Zwislocki and Feldman, 1956). It is not clear whether the decline in interaural phase sensitivity above this frequency is caused by loss of phase-locking in the auditory nerve or to a lower-frequency cutoff particular to the binaural system. Phase-locking at frequencies up to ~ 4 kHz has been observed in the squirrel monkey auditory nerve (Rose *et al.*, 1967), but whether this is representative of the human auditory system is unknown.

In principle, the auditory system should be able to exploit envelope fluctuations in such signals to extract ITD information, but thresholds for detecting ITD changes in amplitude-modulated high-frequency tones are typically much larger than those for low-frequency tones (e.g., McFadden and Pasanen, 1976). Also, several studies have shown that the extent of lateralization possible based on high-frequency envelope ITD is in general small and listener-dependent (Blauert, 1982; Henning, 1974; Trahiotis and Bernstein, 1986). This parallels the typically small, but occasionally substantial, high-frequency ITD weights derived in our experiments I and II. (See Sec. IV D to follow.)

The one aspect in which our results do not correspond to those obtained in lateralization studies is the low weight given to ILD in the low-pass condition of our Experiment I (mean weight, 0.24; see Fig. 5). When presented over headphones without DTF filtering, the intracranial position of low-pass noise is sensitive to ILD. Pinheiro and Tobin (1969) found that the image of broadband noise and noise low-pass filtered at 1.2 kHz could be shifted fully to one side by an ILD of 9 dB. Although naturally occurring ILDs are smallest in the low-frequency regime, this 9-dB value does not exceed the physiological range. For all of our listeners, ILD computed over the 0.5–2-kHz band approached or exceeded 10 dB at lateral angles of 50–60 degrees. Similar ILDs were observed in a 0.8–1-kHz band by Wightman and Kistler (1997a, Fig. 6).

The difference between the low-pass lateralization result and our own might be related to the externalization of our stimuli. Our unbiased lateral angle gain measure indicated that listeners were able to judge accurately the lateral angle of these low-pass targets, and, anecdotally, none reported in-the-head-localization for biased or unbiased low-pass targets. Presumably the 2-oct bandwidth and DTF filtering of these targets was sufficient to create externalized auditory images and perhaps to engage a different mode of cue processing, in which ILD plays a much reduced role in the determination of perceived lateral angle. The role of low-frequency ILDs in near-field distance perception is discussed in Sec. III C 3.

The relative potency of ITD and ILD has been explored in many lateralization studies in which a time-intensity trading ratio was measured. The trading ratio measured using a ''centering'' method describes the amount of time difference favoring one ear that is required to center the image of a signal presented with a level difference favoring the opposite ear. In the ''pointer'' method, the listener adjusts the ITD of a broadband noise to match the lateral location of the experimental stimulus presented with some combination of ITD and ILD (e.g., Moushegian and Jeffress, 1959). Although prone to intersubject differences and level dependencies, a typical result is that the trading ratio for low-frequency stimuli is lower than that for high-frequency stimuli, which indicates that ITD is more potent relative to ILD at low frequencies. For example, Harris (1960) reported trading ratios of ~ 25 μs/dB and ~ 60 μs/dB for low-passed and high-passed clicks, respectively.

Stimuli with images centered using conflicting ITD and ILD cues can be readily discriminated from diotic stimuli (Hafter and Carrier, 1969) perhaps because narrow-band stimuli presented over headphones with conflicting ITD and ILD cues can generate multiple intracranial images (e.g., Whitworth and Jeffress, 1961). The location of the so-called ''time'' image is determined primarily by ITD (and thus exhibits a very low trading ratio), while the ''intensity'' image location is controlled by both ILD and ITD (higher trading ratio). Hafter and Jeffress (1968) found that trading ratios for both types of images were higher for high-passed clicks than

for 500-Hz tone pips, again indicating the reduced potency of ITD at high frequencies. Gaik (1993) examined the conditions under which narrow bands of noise were most likely to produce multiple or noncompact intracranial images. Natural combinations of ITD and ILD in each frequency band were identified using measured directional impulse responses from a human subject. Results showed that the lateralized images were most likely to be unitary and compact when the imposed ITD and ILD were close to a natural pairing of these cues. It is not known whether similar image splitting occurs for wideband noise signals with conflicting binaural difference cues. Our data (see Sec. III C 2) and those of Wightman and Kistler (1992, 1997b) do not provide strong evidence for such an effect in localization experiments.

2. Localization or lateralization?

Although our results are substantially in agreement with many lateralization studies describing the relative strengths of ITD and ILD cues and the perception of binaural stimuli with conflicting time and level cues, we believe that ours was not a lateralization experiment. First, our stimuli consisted of multi-octave, DTF-filtered bands of noise, whereas a majority of trading ratio, binaural discrimination, and extent-of-lateralization studies have used stimuli of restricted bandwidth. Our stimuli thus provided listeners with the opportunity to integrate binaural difference information across frequency, to use monaural spectral cues, and to experience externalized auditory images.

Second, several pieces of evidence make us confident that our listeners perceived externalized (rather than intracranial) images and were localizing them as they would real free-field targets. It was not practical to combine distance estimation with our localization response method, but in a VAS study using ILD-biased stimuli very similar to ours (Wightman and Kistler, 1997b), listeners reported externalized images even for extreme ILD biases. None of our listeners reported trouble making orienting responses to the biased stimuli, which might have been expected were the images not externalized. S18 (the first author), S04 and S93 (laboratory colleagues) all described the images as well externalized and noticed no obvious differences between biased and unbiased stimuli.

Third, our data provide no indication that listeners were responding to multiple images produced by the biased stimuli. Narrow-band signals presented in a lateralization paradigm with conflicting time and amplitude cues often produce multiple intracranial images (i.e., ''time'' and ''intensity'' images). Had this occurred frequently with our VAS stimuli, we would expect to observe bimodality in lateral angle responses and in plots of observed versus imposed bias. This would have been particularly evident in the ILD-biased conditions, for which the location of the ''time'' image would be highly insensitive to the manipulation. Bimodal response patterns were not observed (for example, Fig. 4). Even if multiple images were generated, we are content that our derived weights describe the contributions of ITD and ILD to the image that dominated the percept and drove the listeners' orienting responses. Scatter in listeners' lateral angle responses did not increase markedly as ITD or ILD bias was imposed.

Finally, the strongest evidence that the biased stimuli were localized rather than lateralized comes from the polar angle response data. Imposition of large ITD or ILD biases did produce an increase in front/back confusions for some listeners, but, as in the example shown in Fig. 2, listeners continued to respond accurately to the original elevation of the target. This accuracy would seem unlikely if the orienting responses were derived from a nonexternalized image.

3. Previous localization studies

Our work is not the first to use a localization task in addressing the relative roles of ITD and ILD in spatial hearing. Sandel *et al.* (1955) used a loudspeaker array to produce natural and unnatural combinations of interaural phase and intensity for pure tones, and concluded that ITD was the dominant lateral angle cue for frequencies below 1.5 kHz.

Other researchers have used VAS techniques and noise stimuli to explore the binaural cue weighting. Wightman and Kistler (1992) presented virtual free-field targets in which the interaural phase spectrum was manipulated to correspond to a lateral angle of 90 degrees, rather than to the natural lateral angle of the target. For wideband noise stimuli, listeners' lateral angle judgments agreed with the manipulated interaural phase cue, but with frequencies below 2.5 kHz removed by high-pass filtering, the influence of the fixed ITD cue was almost eliminated for most, but not all, listeners. This result demonstrated the dominance of low-frequency ITD information over other interaural cues and the lack of influence of high-frequency ITD cues. In order to prevent listeners from learning the individual spectral characteristics of their loudspeakers, Wightman and Kistler scrambled the spectra of their noise targets from trial-to-trial in $\frac{1}{3}$-oct bands. This might have obscured or weakened the salience of the veridical spectral cues competing with the manipulated ITD cue. For some of the listeners in our Experiment I, the wideband ITD weight was lower than the low-pass ITD weight, which might indicate increased influence of high-frequency spectral or ILD cues in our unscrambled, wideband noise targets. Wightman and Kistler did not seek to differentiate between the roles of DTF spectra and ILD as the salient cues to lateral angle for high-pass stimuli. The results of our Experiment IV (Sec. VI), however, suggest that DTF spectra have little influence on lateral angle judgments.

Wightman and Kistler (1997b) also obtained results in agreement with ours in a virtual free-field localization condition almost identical to the wideband-ILD condition of our Experiment I. Broadband (0.2–14 kHz) noise targets were presented with an attenuation of 0-40 dB applied to the left-ear signal. For level imbalances (equivalent to our ILD bias) of 10 or even 20 dB, the listeners' responses were generally unperturbed, and corresponded to the position indicated by the unmanipulated ITD and spectral cues. This result was in agreement with our finding that ILD is a weak cue for lateral angle in wideband noise stimuli.

Our results are in agreement with those of Wightman and Kistler (1992, 1997b) that ITD is the dominant lateral

angle cue for stimuli containing low-frequency components. We note, however, that this dominance has been detected only in experiments conducted in anechoic or virtual free-field environments, and might not apply in reverberant environments. Hartmann and Constan (1998) found that the binaural coherence of low-frequency signals presented in rooms is lower than that required to support lateralization on the basis of ITD. Virtual auditory space (VAS) studies incorporating synthesized reverberation would be useful in determining the weighting of cues under nonanechoic conditions.

Low-frequency ILDs for lateral sources in the near-field (<1 m) are much larger than those observed for distant sources, but low-frequency ITDs are not strongly distance-dependent (Brungart and Rabinowitz, 1999). Lateral angle localization judgments remain accurate in the near-field even as low-frequency ILDs diverge from their far-field values (Brungart *et al.*, 1999). These results suggest that, in accord with our finding from the low-pass passband condition of Experiment I, apparent lateral angle for low-pass sources is determined on the basis of ITD (distance-independent) and is independent of ILD (distance-dependent). Low-frequency ILDs have been shown to be important cues for near-field apparent distance (Brungart, 1999). We hypothesize that low-frequency ILDs have little effect on apparent source *direction* because they are *reserved* as cues for near-field *distance* perception.

IV. EXPERIMENT II: WEIGHTING OF ONSET AND ONGOING ENVELOPE-BASED ITD CUES IN HIGH-PASS NOISE

A. Motivation

Experiment I revealed marked individual differences in the perceptual weighting of ITD in high-pass noise stimuli. Because the peripheral auditory system cannot transduce the fine structure of these stimulus waveforms, listeners who placed substantial weight on the high-pass ITD cue must have derived interaural timing information from the envelopes of the signals. Information might be extracted by interaural processing of stimulus onsets or offsets. Moreover, information might be extracted from ongoing fluctuations in the envelopes of the noise waveform, although Middlebrooks and Green (1990) have hypothesized that modulation depth in the outputs of high-frequency auditory filter channels is too small for robust extraction of ongoing ITD cues from noise stimuli.

In experiment II, we explored the relative influence of ITD cues from the onset and ongoing portions of the envelope in high-pass noise stimuli. We attempted to *weaken* the transient envelope ITD cues by *lengthening* the duration of the stimulus onset and offset ramps. In a separate condition, we attempted to *strengthen* the ongoing cues by *amplitude-modulating* the target noise bursts.

B. Stimuli and locations

In Experiment II-A, the high-pass ITD and ILD bias manipulations of Experiment I were repeated with identical sets of target location and bias combinations, but the temporal characteristics of the target noise bursts were altered. The onset and offset ramps of the noise bursts were lengthened from 1 to 20 ms in order to weaken the onset/offset ITD cue. To preserve the 98-ms plateau duration present in the 1-ms-ramped stimuli, the length of the noise bursts was increased to 138 ms. All ten listeners from Experiment I participated in Experiment II-A.

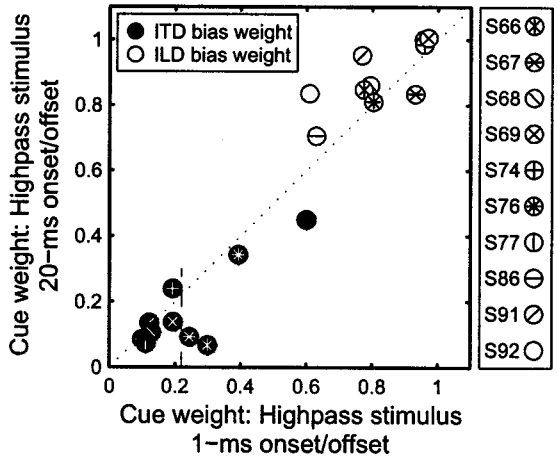


FIG. 6. Effect of onset ramp duration on high-pass ITD and ILD bias weights. The scatter-plot shows high-pass ITD (filled symbols) and ILD (open symbols) bias weights for each of the ten listeners in Experiment II-A. Horizontal axis: cue bias weights for 1-ms onset stimuli; vertical axis: cue bias weights for 20-ms onset stimuli.

In experiment II-B, alterations of the onset/offset and ongoing portions of the envelope were combined factorially. Onset/offset ramps were either short (1 ms) or long (50 ms, 198 ms total duration), and the envelope of the ongoing portion was either *flat* or *modulated*. In the modulated case, the ongoing portion of the signal consisted of 4-ms segments of silence alternating with 6-ms noise bursts with 1-ms onset and offset ramps.[1] The long onsets were intended to reduce the salience of the onset ITD cue, and the amplitude modulation was intended to enhance the salience of the ongoing envelope ITD cue. Data were collected for the high-pass passband condition with either ITD or ILD biases and for all four combinations of the ramp and modulation parameters. Each of the eight stimulus sets (1 passband$\times$2 cue bias types$\times$2 onsets$\times$2 modulations) was presented only once. All other details of the stimulus sets and their presentation were identical to those of Experiment I, including the block size of 111 trials. In the 1-ms-onset/flat-envelope condition for listeners S91 and S92, the data from the equivalent condition of Experiment I were used.

C. Results

1. Experiment II-A

The high-pass ITD and ILD weights for the 20-ms onset stimuli are compared to those for the 1-ms onset targets in Fig. 6. The four listeners whose 1-ms ITD weights were $\geqslant 0.24$ in Experiment I all exhibited reduction of the ITD weight in the 20-ms onset condition. This suggests that they had been relying on the onset ITD cue. The ITD weights for two of these four listeners remained fairly high, however, suggesting either that onset ITD information was still available with the 20-ms onset or that these listeners could exploit

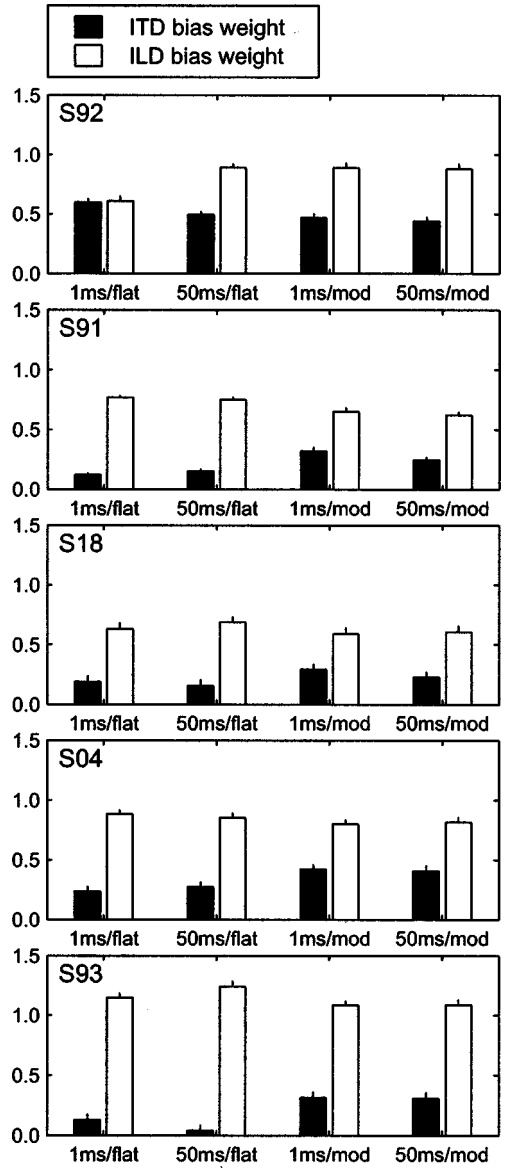


FIG. 7. Effect of onset ramp duration and envelope modulation on high-pass ITD and ILD bias weights. ITD bias weight (black bars) and ILD bias weight (open bars) are plotted for five listeners for each of the four combinations of onset duration (1 or 50 ms) and noise envelope modulation (flat or 100 Hz, 100% modulation depth). Error bars show the standard error of the regression coefficient (i.e., of the cue bias weight).

intrinsic envelope fluctuations to obtain ongoing ITD information. Further evidence for the reduced salience of the onset ITD in the 20-ms condition is provided by the increase in the ILD weights obtained for the longer-onset stimuli. In ILD bias conditions, the ITD cue was consistent with the unbiased target location. Thus weakening of the ITD cue should have reduced the influence of cues competing with the biased ILD.

2. Experiment II-B

ITD and ILD weights for the stimuli of Experiment II-B are shown for five listeners in Fig. 7. The unbiased lateral angle gain was consistent across envelope condition for all listeners, which indicates that the envelope manipulations did not affect the listeners' localization accuracy. In the 1-ms-onset/flat-envelope condition, the weights for listeners S18, S04, and S93 (0.19, 0.24, and 0.13, respectively) were consistent with the distribution of weights observed in the high-pass ITD bias condition of Experiment I.

We consider first the effect of the onset-time manipulation on ITD weight. In the flat-envelope conditions, the effect of lengthening the signal onset and offset from 1 to 50 ms was small (mean ITD weight reduction, 0.03) except for listener S92, whose larger ITD weight decreased from 0.60 to 0.50. Thus, as found in experiment II-A, weakening the onset ITD cue had little impact on the listeners whose ITD bias weights were already small in the 1-ms-onset condition (S91, S18, S04, and S93), likely because of a simple floor effect. For S91 and S04, the ITD bias weights actually increased slightly (by only 0.03 and 0.04, respectively). In the modulated-envelope conditions, lengthening the onset had little effect on the ITD weight for any listener (mean ITD weight reduction, 0.04).

In contrast, the envelope modulation manipulation had a pronounced effect in both the 1- and 50-ms onset conditions. Four of the five listeners exhibited an increase in the weighting of the ITD cue in both onset-length conditions when envelope modulation was added (for these four, mean ITD weight increase over both onset lengths, 0.16). For these four listeners, adding envelope modulation also produced modest decreases in the ILD weight (mean ILD weight reduction, 0.09). A similar inverse relation between ITD and ILD weights was observed in Experiment II-A. Listener S92, for whom the high-pass ITD weight in Experiment I was already high (0.60), maintained a consistently high weight on the ITD cue but did not exhibit an increase in the weight in the modulated-envelope conditions.

Together, the results of Experiments II-A and II-B suggest that both onset and ongoing envelope ITD cues play a role in the sensitivity of some listeners to high-frequency ITD. It appears, however, that the highest high-frequency ITD bias weights are obtained when the listener is able to process ongoing envelope ITD cues, and that under such circumstances, the onset cue is of reduced importance.

D. Discussion: Role of high-frequency envelope ITD cues

As noted previously, complex, high-frequency stimuli such as amplitude-modulated tones, tone complexes, and bands of noise provide the auditory system with ongoing envelope ITD information that is absent from high-frequency pure-tone stimuli. ITD discrimination thresholds measured for such sounds are typically found to be two to ten times larger than those for low-frequency tones—as large as a few hundred μs (e.g., Bernstein and Trahiotis, 1994; Blauert, 1982; Henning, 1974; McFadden and Pasanen, 1976; Nuetzel and Hafter, 1981). Although easily detectable, ITDs of physiologically plausible magnitude have been found to be remarkably ineffective in displacing the intracranial images of high-frequency, complex stimuli away from the midline in lateralization experiments, and both detection thresholds and lateral displacement sensitivity display marked individual differences (Trahiotis and Bernstein, 1986).

Our results from Experiment I paralleled these findings (in a localization rather than lateralization task) both with respect to the generally low weights accorded to high-frequency ITD and to the variation of those weights among listeners.

The results of Experiments II-A and II-B show that (a) some listeners with substantial weights on high-frequency ITD derive this information primarily from the onset ITD cue, (b) others are able to extract ongoing ITD from intrinsic noise envelope fluctuations, and (c) if robust envelope modulation is present, weights on high-frequency ITD increase and the strength of the onset cue has little effect. These results are consistent with the findings that listeners vary in their ability to extract envelope-based, ongoing ITD information, and that onset cues are salient only when the normally potent ongoing cues are ambiguous. The psychophysical literature contains a diversity of reports on the relative strength and discriminability of onset and ongoing ITD cues (for example, Buell *et al.*, 1991; Hafter and Dye, 1983; Tobias and Schubert, 1959). In a synthesis of these and other results, Freyman and colleagues (1997) concluded that results differ because the contribution of the onset cue is strongly dependent upon the spectral, temporal, and binaural characteristics of the ongoing stimulus. In particular, ''lateralization of a spectrally dense signal with an unambiguous ongoing delay is not subject to dominance by onsets even if the onset cue itself is strong'' (Freyman *et al.*, 1997). In our Experiment II, the high-pass noise signals satisfied the condition of spectral density, and ambiguity in ongoing envelope ITDs was reduced by applying envelope modulation.

The results of a simulation of high-frequency, envelope-based ITD discrimination (see Appendix) suggest that intrinsic envelope fluctuations in unmodulated high-frequency noise bands are not sufficiently large for useful discrimination of ongoing, envelope-based ITDs. This result supports the hypothesis of Middlebrooks and Green (1990). Modulation depth was found to be independent of differences in listeners' DTFs, and therefore individual differences in high-pass ITD bias weights must be related to individual differences in envelope extraction processes or in information processing strategies. Constan and Hartmann (2001) have shown that in reverberant environments such as those in typical rooms, the coherence of the two ear signals is often insufficient to permit lateralization on the basis of high-frequency envelope ITD. This may be another reason that most listeners discount these cues.

For noise stimuli, intrinsic envelope fluctuations in different auditory filter channels are not correlated. That is, the channel signals are not *comodulated*. Comodulation is known to be an important factor in across-frequency integration of information (e.g., Hall and Grose, 1990), and it is possible that, for noise stimuli, the lack of comodulation among envelopes in different auditory filter bands reduces the salience of the high-frequency envelope cue for most listeners. The increased ITD bias weight we observed when amplitude modulation was imposed on the noise targets in Experiment II-B might have been a function of increased interchannel comodulation as well as increased modulation depth. Saberi (1995) has shown that ITD discrimination thresholds for stimuli composed of two spectrally separated, high-frequency, narrow bands of noise are lower when the temporal envelopes of the bands are identical than when they are different. Similarly (but in a low-frequency regime), Trahiotis and Stern (1994) have found that complexes of spectrally separated SAM tones with consistent carrier ITDs produce compact unitary binaural images only when identical modulators are applied to each carrier.

Eberle *et al.* (2000) investigated the salience of envelope-based ITD cues in a free-field localization task. Listeners reported the apparent location of a high-frequency octave band of noise (7–14 kHz), with and without an applied 20- 80- or 320-Hz, 100%, sinusoidal amplitude modulation. This signal was similar to the modulated-envelope targets used in Experiment II-B of the present study. The authors found that the introduction of amplitude modulation produced no reduction in the mean magnitude of errors in listeners' lateral angle judgements. They inferred that the amplitude modulation did not facilitate the extraction of envelope-based, high-frequency ITD cues. The results of our Experiment II-B suggest that ITD salience is enhanced by amplitude modulation, and thus are not in agreement with this conclusion.

The discrepancy between our conclusions and those of Eberle *et al.* (2000) likely resulted from the fact that our VAS technique permitted the elimination of a confound between ITD and ILD cues inevitably present in the free-field situation. In our Experiments I and II, listeners were able to judge accurately the lateral angle of the unmanipulated high-pass targets despite their generally low sensitivity to the ITD cue. The veridical judgements were, therefore, likely to have been based on the ILD or on spectral cues (but see Sec. VI D to follow). Using free-field target presentation, it is not possible to present stimuli in which the various localization cues are in conflict. If, similarly to our subjects, the listeners in the Eberle *et al.* (2000) study were able to judge lateral angle accurately on the basis of the sufficient and veridical ILD cues, then adding ITD information consistent with the ILD via amplitude modulation would not have changed the apparent location of the source. The consistently low lateral localization error across conditions might reflect the resolution of the motor response method rather than evidence of complete discounting of the high-frequency ITD cue.

V. EXPERIMENT III: WEIGHTING OF THE INTERAURAL LEVEL SPECTRUM CUE

A. Motivation

The frequency-independent ILD bias manipulation used in Experiments I and II resulted in stimuli with unnatural *interaural level spectra* (ILS; i.e., patterns of ILD across frequency). As a wideband sound source is moved away from the median plane, ILD grows more rapidly at high frequencies than at low ones, and thus the frequency-independent ILD created by applying ILD bias to a midline source in Experiments I and II would never be observed under natural conditions. In principle, the ILS could serve as a robust cue to source elevation and front-back location for wideband sources (Duda, 1997), because it is the difference

between the spectra at the left and right ears, and therefore largely independent of irregularities in the source spectrum. There is, however, little psychophysical evidence to support the salience of ILS as a vertical-plane localization cue (see Discussion, Sec. V E).

In Experiment III, we attempted to bias the perceived location of virtual free-field targets without introducing unnatural patterns of ILD across the frequency spectrum.

B. Stimuli and locations

The targets were 100-ms noise bursts with 1-ms raised-cosine onsets and offsets. Stimuli with biased interaural level spectra were generated as follows. Starting with an original target location, **A**, we identified a second location, **B**, with the same polar angle, but displaced in lateral angle from **A** by ± 30 or ± 60 degrees.

For the ear on the same side as the target (the *near* ear), we retained the DTF of location **A**, $\mathrm{DTF}_{\mathrm{near}}(\mathbf{A},f)$. For the opposite ear (the *far* ear), we generated a new transfer function, $X(f)$, such that the resulting ILS matched that of location **B**. That is,

$$\left|\frac{\mathrm{DTF}_{\mathrm{near}}(\mathbf{A},f)}{X(f)}\right| = \mathrm{ILS}(\mathbf{B},f) = \left|\frac{\mathrm{DTF}_{\mathrm{near}}(\mathbf{B},f)}{\mathrm{DTF}_{\mathrm{far}}(\mathbf{B},f)}\right|.$$

It follows that

$$|X(f)| = \left|\mathrm{DTF}_{\mathrm{near}}(\mathbf{A},f)\times\frac{\mathrm{DTF}_{\mathrm{far}}(\mathbf{B},f)}{\mathrm{DTF}_{\mathrm{near}}(\mathbf{B},f)}\right|.$$

In practice, we computed $X(f)=\mathrm{DTF}_{\mathrm{far}}(\mathbf{A},f)\times H(f)$, where $H(f)$ was the *zero-phase* filter with transfer function

$$H(f) = \left|\frac{\mathrm{DTF}_{\mathrm{near}}(\mathbf{A},f)}{\mathrm{DTF}_{\mathrm{far}}(\mathbf{A},f)}\right| \Big/ \left|\frac{\mathrm{DTF}_{\mathrm{near}}(\mathbf{B},f)}{\mathrm{DTF}_{\mathrm{far}}(\mathbf{B},f)}\right| = \frac{\mathrm{ILS}(\mathbf{A},f)}{\mathrm{ILS}(\mathbf{B},f)}.$$

This operation preserved the natural near-ear DTF and the interaural phase spectrum (and hence the ITD) measured at location **A**, but altered the ILS to correspond to that of location **B** (see Table I). Positive (rightward) ILS bias was applied only to targets originally on or to the right of the median plane. Similarly, negative ILS bias was applied only to midline or left-hemisphere targets. Thus in the above expression for $H(f)$, the ear designated ''near'' remained the same for location **A** and location **B**, as did the ear designated ''far.''

The wideband, low-pass, and high-pass passband conditions used in Experiment I were employed. The groups of target locations for Experiment III were similar to those for Experiments I and II, but because ILS bias was applied only to shift apparent position laterally *away* from the median plane, the total number of biased targets was reduced from 144 to 80. In total there were three stimulus sets (3 passbands$\times$1 cue bias type) of 158 targets each. The targets in each set were presented twice in a shuffled order over the course of four blocks of 79 trials each. Blocks from different passband conditions were intermixed within an experimental session but trials from different sets were not intermixed within blocks.

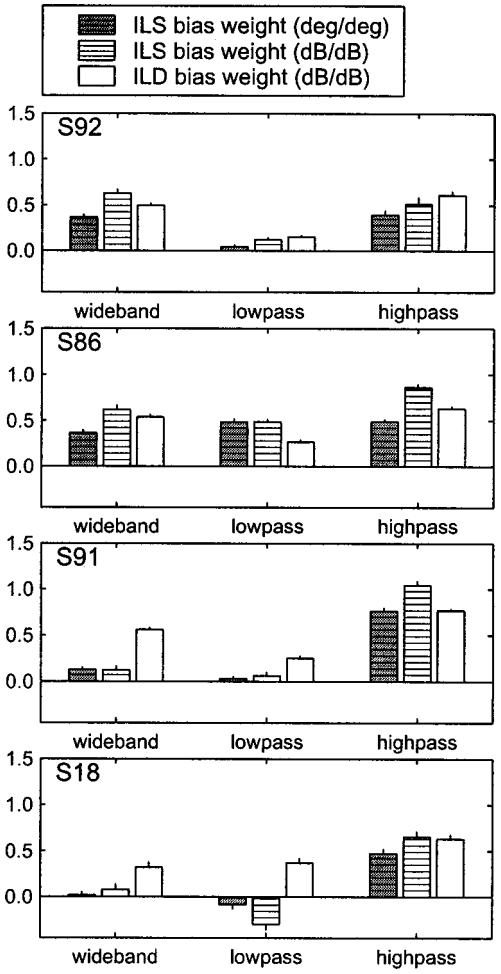


FIG. 8. Interaural level spectrum and interaural level difference bias weights for Experiment III. Each panel shows data for one listener. ILS-degree/degree bias weight (shaded/hatched bars), ILS-dB/dB bias weight (unshaded/hatched bars) and ILD bias weight (open bars) are plotted for the wideband, low-pass and high-pass passband conditions. Error bars show the standard error of the regression coefficient (i.e., of the cue bias weight).

C. Analysis

Because the change in the ILS caused by the manipulation was not a scalar quantity, as were the imposed ITD or ILD biases of Experiments I and II, we first computed a weight for the ILS cue by simply computing the slope of a linear fit to the observed lateral angle response bias versus imposed ILS bias data, both measured in degrees. This gave a weight in units of degree/degree. We also analyzed the responses as we did in Experiment I, treating *gross ILD*, the overall interaural difference in energy integrated across the stimulus passband, as the effective cue. In this case, the imposed bias was computed from the difference between the interaural energy ratios present in the original and modified DTFs, and the observed bias was computed as in Experiment I (Sec. II E). A linear regression between the observed and imposed biases yielded an ILS bias weight in units of dB/dB.

D. Results

The cue weights derived from the responses to ILS-biased stimuli are shown in Fig. 8, along with the corresponding ILD weights from Experiment I (or equivalent con-

ditions for S18). In general the relation between ILS-degree/degree weights (shaded/hatched bars) and passband condition was similar to that found for the ILD weights (open bars) of Experiment I. That is, low weight was given to the ILS cue in the low-pass condition (three of four listeners), and a higher weight was given in the high-pass condition. Curiously, small negative ILS-degree/degree and ILS-dB/dB weights were observed for listener S18 in the low-pass condition.

Several features of the data suggest that the detailed shape of the interaural level spectrum itself is not a particularly salient cue for lateral angle, but rather that the effective interaural level cue is the overall energy difference between the ears. For three of the four listeners, the ILS-degree/degree weight was lower than the ILD weight from experiment I in both the wideband and high-pass conditions. Also, in the majority of cases in the wideband and high-pass conditions, the ILS-dB/dB (gross ILD) weight was higher than the ILS-degree/degree weight computed from the same responses.

In some cases, a particular gross ILD produced more response bias when it was produced by the ILS manipulation than when it was applied as a flat attenuation. Evidence for this can be found in the high-pass weights for S86 and S91. For these listeners, the high-pass ILS-dB/dB weight (Fig. 8, hatched bars) was substantially higher than the high-pass ILD weight from Experiment I (open bars).

E. Discussion: Interaural level spectrum and the nature of ILD processing

The shape of the interaural level spectrum has been proposed as a cue for vertical-plane localization for locations both on (Searle *et al.*, 1975) and off (Duda, 1997) the median plane. The results of several psychophysical studies contradict this proposal. First, vertical-plane localization of wideband sources can be disrupted by certain source-spectrum irregularities which preserve the ILS (e.g., Macpherson, 1996, 1998; Rakerd *et al.*, 1999; Wightman and Kistler, 1997a). Second, there is evidence that the shape of the spectrum at the ear contralateral to the target can be altered substantially without consequent degradation in vertical-plane localization accuracy (Humanski and Butler, 1988; Morimoto, 2001; Wightman and Kistler, 1999), and that the influence of the contralateral-ear DTF declines as the source location is moved away from the median plane (Morimoto, 2001). These results suggest that the details of the ILS are unimportant for human listeners' vertical-plane localization, and our results from Experiment III lead to a similar conclusion about the lack of efficacy of the ILS as a cue to lateral angle.

The effective cue for lateral angle in high-pass stimuli seems to be better modeled as the overall difference in proximal stimulus energy between the ears, or, perhaps more realistically, as an integration of the ILDs observed in discrete frequency channels. This would parallel the auditory system's use of low-frequency ITD as a cue to lateral angle and its apparent insensitivity to fine details in the interaural phase spectrum (Kulkarni *et al.*, 1999). Both the ITD and ILD findings are consistent with the suggestion that " . . . the system does not evaluate every detail of the complicated interaural dissimilarities, but rather derives what information is needed from definite, easily recognizable attributes" (Blauert, 1997, p. 138).

The results of Experiment III do, however, reveal some effect of the distribution of ILD across frequency. As discussed earlier, we observed that a net ILD bias generated by the ILS manipulation was more effective in biasing lateral angle judgments than was an equivalent flat interaural attenuation (Fig. 8). That is, the dB/dB weights for ILD obtained in Experiment III were higher than those obtained in Experiment I. A possible explanation for this is that listeners placed greater weight on the gross ILD when the ILS was more natural, or at least when it did not contain unnatural low-frequency ILDs. This would be consistent with the proposal of Wightman and Kistler (1997a) that the naturalness of observed localization cues plays a mediating role in their salience.

VI. EXPERIMENT IV: WEIGHTING OF THE NEAR-EAR SPECTRAL CUE

A. Motivation

Some have suggested that monaural spectral cues are important in determining perceived lateral angle as well as being the primary cues to sound source elevation and front/back location (for example Butler and Flannery, 1980). In Experiments I, II, and III of the present study, we could not directly assess the influence of spectral cues on perceived lateral angle because we did not manipulate these cues independently of both of the binaural difference cues (ITD and ILD) simultaneously. In all cases, the spectral cue in the ear on the side of the original target location corresponded with the unmanipulated binaural difference cue. The results of Experiments I and II suggest, however, that the spectral cue contribution is minimal, as we discuss below in Sec. VI D.

In Experiment IV, we biased the lateral angle corresponding to the spectral cue in one ear independently of both the ITD and ILD cues. The near-ear DTF (as defined in Sec. V B) was replaced by another corresponding to a new location displaced in lateral angle from the original target location. The opposite-ear impulse response was altered in order to maintain the original interaural level and phase spectra. We manipulated the near-ear DTF because it is thought to be the most potent cue for vertical plane localization and because, as discussed earlier, it appears that the far-ear spectrum can be manipulated severely without affecting vertical plane localization (Humanski and Butler, 1988; Morimoto, 2001; Wightman and Kistler, 1999).

B. Stimuli and locations

As in the previous experiments, the targets were 100-ms noise bursts with 1-ms raised-cosine onsets and offsets. Stimuli with biased near-ear DTFs were generated as follows. Starting with an original target location, **A**, we identified a second location, **B**, with the same polar angle, but displaced in lateral angle from **A** by ± 30 or ± 60 degrees.

The interaural level- and phase-difference spectra for location **A** were computed as

$$\mathrm{ILS}(\mathbf{A},f)=\left|\frac{\mathrm{DTF}_{\mathrm{near}}(\mathbf{A},f)}{\mathrm{DTF}_{\mathrm{far}}(\mathbf{A},f)}\right|$$

and

$$\mathrm{IPS}(\mathbf{A},f)=\phi\{\mathrm{DTF}_{\mathrm{near}}(\mathbf{A},f)\}-\phi\{\mathrm{DTF}_{\mathrm{far}}(\mathbf{A},f)\}.$$

For the ear on the target side, $\mathrm{DTF}_{\mathrm{near}}(\mathbf{B},f)$ was substituted for $\mathrm{DTF}_{\mathrm{near}}(\mathbf{A},f)$; this constituted the spectral cue bias. In the far-ear channel, a new transfer function, $X(f)$, was synthesized with magnitude

$$|X(f)|=\left|\frac{\mathrm{DTF}_{\mathrm{near}}(\mathbf{B},f)}{\mathrm{ILS}(\mathbf{A},f)}\right|$$

and phase

$$\phi\{X(f)\}=\phi\{\mathrm{DTF}_{\mathrm{near}}(\mathbf{A},f)\}-IPS(\mathbf{A},f).$$

Thus the stimulus had the ITD and ILS cues of location **A**, but the near-ear DTF of location **B**. This differed from experiment III, in which the near-ear DTF corresponded to **A** and the ILS to **B** (see Table I). Prior to computing the far-ear DTF magnitude spectrum, $\mathrm{ILS}(\mathbf{A},f)$ was smoothed by frequency-domain convolution with a 1-kHz-wide rectangular filter. This reduced the creation of sharp resonances in the synthesized DTF caused by notches in $\mathrm{DTF}_{\mathrm{near}}(\mathbf{A},f)$.

Biased targets were located at lateral angles of ± 30 or ± 60 degrees and at elevations of 0, ± 20, and ± 40 degrees in the front and rear hemispheres. At each location there were three possible biases that did not place locations **A** and **B** on opposite sides of the median plane. For example, a target at a lateral angle of -30 degrees could be biased by -60, -30, or $+30$ degrees without crossing the median plane.

The wideband and high-pass passband conditions used in Experiment I were employed. In total there were two stimulus sets (2 passbands$\times$1 cue bias type) of 198 targets each. Of these targets, 120 were filtered and biased, 40 were filtered but unbiased, and the remaining 38 were wideband, unmanipulated targets. The targets in each set were presented once in a shuffled order over the course of two blocks of 99 trials each. Three listeners (S04, S18, S93) who had participated in Experiment II-B participated in Experiment IV.

C. Results

We computed a bias weight for the ispilateral DTF cue by simply computing the slope of a linear fit to the observed lateral angle response bias data and the imposed DTF bias data, both measured in degrees. This gave a weight in units of degree/degree, as in our first analysis of the ILS bias data in Experiment III (Sec. V C). Although our manipulation preserved the ILS of the original target location, it did not necessarily preserve the overall ILD because changing the near-ear spectrum changed the spectral distribution of energy. Because we did not wish our results to be confounded by an unintended ILD bias, trials for which the overall stimulus ILD differed from that of the original target location by more than 5 dB were excluded from the analysis. The proportion of excluded trials varied between 4% and 10% among listeners.

The obtained weights are shown in Table II. For all listeners and in both passband conditions, the magnitudes of all weights were <0.1, and in a majority of cases were not significantly different from 0. These results suggest that monaural spectral cues play a negligible role in determining apparent lateral angle even in the high-pass condition, in which the strong conflicting low-frequency ITD cue was removed.

TABLE II. Measured DTF-bias weights for the wideband and high-pass targets of Experiment IV.

Listener	Wideband	High-pass
S04	-0.01 ± 0.03	0.09 ± 0.04
S18	-0.05 ± 0.05	0.00 ± 0.06
S93	0.08 ± 0.04	0.05 ± 0.05

D. Discussion: Influence of spectral cues on lateral localization

In Experiments I and II of the present study, we did not manipulate monaural spectral cues independently of ITD and ILD to assess their influence on perceived lateral angle, but those results also suggest that the spectral cue contribution is minimal. In each stimulus condition of these experiments, the spectral cues at both ears corresponded to the original target location prior to the ITD or ILD manipulation and were in agreement with the unmanipulated interaural cue. Thus, any reliance on the spectral cues should have resulted in a reduction of the weight placed on the manipulated cue, although, considering each condition in isolation, the effects of spectral cues and the unmanipulated cue cannot be disambiguated. Our finding, however, that weights approaching unity were found for one or other of the interaural cues in each passband condition is evidence that spectral cues do not contribute substantially to lateral angle localization even in a binaural listening situation. A high weight on a biased interaural cue indicates that it was possible to shift the apparent position of the target to a location inconsistent with the natural spectral cues. In Experiment IV, we directly tested the influence of monaural spectral cues by biasing the source-side DTF independently of the natural ITD and ILS cues. We found that the DTF bias had little or no influence on the perceived lateral angle of the target.

These results are in agreement with and extend the work of Slattery and Middlebrooks (1994) and of Wightman and Kistler (1997b), who found that normal-hearing listeners rendered monaural by plugging one ear were incapable of accurate lateral angle judgements. In such an acute monaural situation, listeners reported all localization targets originating from a position opposite the unoccluded ear. Slattery and Middlebrooks did find that approximately half of their chronically monaural listeners were able to make use of spectral information to determine source azimuth, but they considered this to be the result of a long-term adaptation to the listeners' lack of access to binaural cues, rather than being indicative of the functioning of the normal auditory system.

VII. SUMMARY AND CONCLUSIONS

Our results suggest that, in broad outline, the duplex theory does serve as a useful description of (if not a principled explanation for) the relative potency of ITD and ILD cues in low- and high-frequency regimes. In Experiment I, we found that listeners weighted the ITD cue strongly as a cue for lateral angle for low-pass stimuli and (with some exceptions) weighted ITD weakly for high-pass stimuli. The opposite pattern was observed for listeners' ILD weights. Even when substantial biases were introduced, ILDs were generally ignored at low frequencies but were given high weights for high-pass stimuli. For wideband targets, both cues were given substantial weight, but ITD dominated for most listeners. In Experiment II, we found that both onset and ongoing envelope ITD cues contributed to the relatively minor role of high-frequency ITD information, but that the greatest high-frequency ITD weights were observed for listeners who were able to make use of ongoing cues. In Experiment III, we examined the role of the detailed shape of the interaural level spectrum as a lateral angle cue. The results suggested that the precise shape of the ILS was not as effective a cue as the overall energy difference between the ears. In Experiment IV, we found that monaural spectral cues had little or no influence on perceived lateral angle.

ACKNOWLEDGMENTS

The authors are grateful to Leslie Bernstein, Fred Wightman, and an anonymous reviewer for their constructive comments on an earlier version of this paper. William Hartmann, Zachary Constan, Douglas Brungart, Brian Mickey, and Christopher Stecker also provided helpful suggestions. Zekiye Onsan provided invaluable technical assistance. This work was funded by NIH Grant Nos. R01DC00420 and T32DC00011.

APPENDIX: CAN INTRINSIC ENVELOPE FLUCTUATIONS IN WIDEBAND NOISE SUPPORT ITD DISCRIMINATION AT HIGH FREQUENCIES?

1. Motivation

For most of our listeners in experiment I, ITD had very little influence on judgments of lateral angle for high-pass noise stimuli. One possible explanation for this result is that the auditory system is unable to extract robust ITD information from such signals, as suggested by Middlebrooks and Green (1990). Alternatively, the system might in principle be able to obtain useful ITD information, but listeners discount that information for some other reason. Psychophysical studies of ITD sensitivity in high-frequency noise have tended to use narrowband stimuli (e.g., Bernstein and Trahiotis, 1994), which might encourage listeners to adopt an off-frequency listening strategy not possible with a broadband stimulus. Because psychophysical data are not available for the multi-octave high-pass stimuli we used in Experiments I and II, we conducted a series of simulations in order to determine whether high-frequency, envelope-based ITD is indeed a viable lateral angle cue for wideband noise stimuli.

To play a role in high-frequency ITD sensitivity, modulations in the right- and left-ear signal envelopes must exhibit two properties. *First*, the envelope fluctuations must be highly *coherent* (i.e., similar), or else ITD cannot be defined. For wideband noise sources off the midline, the strong asymmetry between source-side and far-side DTFs leads to decorrelation of the wideband right- and left-ear envelopes. This occurs because the envelopes of the narrowband components that compose the wideband noise are uncorrelated and the overall envelope depends on the relative levels (and phases) of these components. Within individual auditory filter bands, however, this decorrelation might not occur if the effect of interaural DTF asymmetry is primarily the imposition of a level difference, which would not alter the shapes of the envelopes.

Second, the envelope modulations must be of sufficient *depth*. In a study of interaural envelope delays, Middlebrooks and Green (1990) hypothesized that, because of the low-pass limits of high-frequency envelope following, the depth of intrinsic noise envelope modulations in high-frequency auditory filter bands might be too small to allow effective extraction of envelope-based ITD information from noise signals. Reduction of modulation depth increases ITD just-noticeable differences (jnd's; i.e., discrimination thresholds) in sinusoidally amplitude-modulated (SAM) tones and beating two-tone complexes (Henning, 1974; McFadden and Pasanen, 1976; Nuetzel and Hafter, 1981).

A metric which simultaneously captures the effects of changes in modulation depth and interaural envelope coherence is the *normalized correlation* of the left- and right-ear signal envelopes. This is computed in the same manner as the Pearson product-moment correlation (or *normalized covariance*), but the mean (or d.c.) components of the envelopes are retained. Using this metric, Bernstein and Trahiotis (1996) have successfully accounted for the dependence of ITD jnd's on both SAM-tone modulation depth and two-tone modulation depth as measured by Nuetzel and Hafter (1981) and Pasanen (1976), respectively.

In our simulations, we passed wideband Gaussian noise signals through an auditory filter-bank model. We then estimated ITD discrimination thresholds (Δ_{ITD}) based on the outputs of individual filter channels by determining the minimum increment in ITD required to reduce the normalized correlation by a criterion amount. We derived estimates of minimum audible angle (MAAs) from these Δ_{ITD} estimates. We also examined the sensitivity of the jnd estimates to DTF filtering of the noise signals and to the low-pass cutoff frequency of the envelope-following process.

2. Methods

We passed 16 100-ms exemplars of wideband Gaussian noise through a gammatone auditory filter-bank model (Slaney, 1994) with binaural channels centered at 4, 6, 8, 10, 12, and 14 kHz. Envelopes for each ear in each frequency band were extracted by half-wave rectification followed by low-pass filtering at either 250 or 500 Hz (fourth-order Butterworth filter). All processing was done at a sampling rate of 50 kHz.

To estimate an ITD discrimination threshold for each filter channel and low-pass filter cutoff, we first computed the normalized correlation, ρ, of the right- and left-channel envelopes as we imposed increasing interaural delays in 20-μs (1-sample) steps. The resulting cross-correlation function had a peak value of 1 at 0-lag, and declined monotonically with increasing lag. The imposed interaural delay required to reduce the normalized correlation to 0.99 was then taken as the estimate of the ITD jnd. This criterion matched the mean threshold decrement in normalized correlation of $\sim$0.01 found by Bernstein and Trahiotis (1996, Fig. 4) for their four most sensitive listeners. For each filter channel and low-pass filter cutoff, the final jnd estimate was taken as the mean value over the 16 samples of Gaussian noise.

As a descriptive statistic, we also estimated the envelope modulation depth in each filter channel for each low-pass filter cutoff. Envelope modulation depth is a measure of the magnitude of a signal's amplitude fluctuations relative to the mean level of the signal. For a sinusoidally amplitude-modulated tone, the modulation depth, m_{SAM}, is defined as the ratio of the modulation amplitude, a (half the peak-to-trough range) and the mean level:

$$m_{\mathrm{SAM}}=\frac{a}{\mathrm{mean}_{\mathrm{ENV}}}=\frac{\frac{1}{2}(\max_{\mathrm{ENV}}-\min_{\mathrm{ENV}})}{\frac{1}{2}(\max_{\mathrm{ENV}}+\min_{\mathrm{ENV}})}.$$

For sinusoidal modulation, $a=\sqrt{2}\sigma_{\mathrm{SAM}}$, where σ is the standard (or root-mean-square) deviation of the envelope about the mean level. Using this relation we derived the more general expression

$$m=\frac{\sqrt{2}\sigma_{\mathrm{ENV}}}{\mu_{\mathrm{ENV}}}$$

(in which σ_{ENV} and μ_{ENV} are respectively the estimated standard deviation and mean of the envelope) as an estimator for the modulation depth of a nonsinusoidal envelope.

To investigate how DTF filtering might affect ITD discrimination, for each of the participants in Experiments I–III, we also filtered samples of Gaussian noise with the measured right- and left-ear DTFs corresponding to azimuths of 0, 30, 60, and 90 degrees in the horizontal plane. For each listener, azimuth, filterbank channel and low-pass filter cutoff, we repeated the ITD jnd and modulation-depth estimation procedures using these DTF-filtered signals. Before computing Δ_{ITD}, we removed the DTF-related ITD by time-aligning the right- and left-ear envelopes on a channel-by-channel basis to maximize the normalized correlation. The peak value of ρ was recorded and the jnd was then computed relative to the adjusted ITD. Psychophysical data are not available for discrimination of changes in envelope correlation for pedestal correlations other than 1, but in the majority of cases the peak value of normalized correlation observed was $>$0.95, and we therefore retained the 0.01 reduction in ρ as a reasonable threshold criterion.

3. Results and discussion

We first consider the results obtained without DTF filtering. Estimates of Δ_{ITD} with the 500-Hz low-pass envelope filtering increased roughly linearly with increasing channel frequency from a mean of 256 μs at 4 kHz to 298 μs at 14 kHz. This range of thresholds is similar to those measured by Bernstein and Trahiotis (1994) for noises of various bandwidths centered at 4 and 8 kHz, for which the mean jnd's were $\sim$150 and $\sim$300 μs, respectively. Our jnd estimates were very sensitive to the cutoff frequency of the low-pass filter used to extract the envelope; using the 250-Hz cutoff, Δ_{ITD} increased to 511 μs at 4 kHz and to 740 μs at 14 kHz. These values are substantially higher than those found psychophysically with bands of noise, suggesting that the 250-Hz low-pass filter cutoff was too low.

In the absence of DTF filtering, the right- and left-ear envelopes were perfectly correlated, and therefore the reduction in estimated ITD acuity must have been caused by reduction in modulation depth, m. With 500-Hz low-pass filtering, m declined approximately linearly with increasing channel frequency from a mean of 0.65 (i.e., 65% modulation) in the 4-kHz channel to 0.43 in the 14-kHz channel. Reduction of the low-pass cutoff to 250 Hz lowered modulation depth to 0.53 at 4 kHz and to 0.34 at 14 kHz. These dependencies on frequency and low-pass cutoff were expected because of the increase of channel bandwidth with increasing center frequency. The high-frequency limit of channel envelope fluctuations is roughly proportional to channel bandwidth (Dau *et al.*, 1999; Lawson and Uhlenbeck, 1950), and therefore a fixed low-pass filter cutoff removes a greater proportion of the envelope fluctuation energy in the higher-frequency channels.

The addition of DTF filtering caused an increase in Δ_{ITD} of 50–100 μs for most listeners and most frequency channels. There was no consistent effect on modulation depth of listener, azimuth, or near/far ear, and therefore the loss of sensitivity must have been caused by decorrelation of the envelopes related to DTF asymmetry. This manifested itself in reduction of the peak value of the normalized correlation, which, however, remained $>$0.95 unless a pronounced spectral notch in either the left or right DTF was aligned with the center frequency of the filter channel. In such cases, the peak normalized correlation was further reduced, and typically fell in the range 0.90–0.95, presumably because of decorrelation of the right- and left-channel envelopes caused by the rapid changes in phase associated with the spectral notches. The consequence of this decorrelation was to increase Δ_{ITD} in the affected frequency bands by at least 200–300 μs.

An ITD jnd on the order of 250 μs corresponds to an MAA of approximately 25 degrees because ITD varies by $\sim$10 μs per degree of lateral angle across the midline. Although one would expect that combining ITD information across frequency channels would lead to lower overall thresholds, such large MAAs do not appear small enough to permit accurate lateral localization. These simulation results suggest that, in agreement with the hypothesis of Middlebrooks and Green (1990), high-frequency Gaussian noise signal envelopes *cannot* convey robust ITD information.

[1]In effect, the ongoing portion of the noise burst was amplitude modulated at a depth of 100% by a 100-Hz, smoothed square-wave modulator. We computed the levels of the low-frequency sidebands generated by this modulator, and determined that any sidebands appearing below 1.5 kHz were at least 60 dB lower in level than components of the high-frequency passband. Bernstein and Trahiotis (1982) caution that low-frequency sidebands as low in level as 50 dB below the passband components of nominally high-pass signals can be informative for ITD discrimination. We believe, and the small ITD weights observed in the envelope-modulated conditions suggest, that the sidebands in our stimuli were low enough to avoid this problem.

Bernstein, L. R., and Trahiotis, C. (**1982**). ''Detection of interaural delay in high-frequency noise,'' J. Acoust. Soc. Am. **71**, 147–152.

Bernstein, L. R., and Trahiotis, C. (**1994**). ''Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise,'' J. Acoust. Soc. Am. **95**, 3561–3567.

Bernstein, L. R., and Trahiotis, C. (**1996**). ''On the use of the normalized correlation as an index of interaural envelope correlation,'' J. Acoust. Soc. Am. **100**, 1754–1763.

Blauert, J. (**1982**). ''Binaural localization: Multiple images and applications in room and electroacoustics,'' in *Localization of Sound: Theory and Applications*, edited by R. W. Gatehouse (Amphora, Groton, CT), pp. 3–25.

Blauert, J. (**1997**). *Spatial Hearing* (MIT, Cambridge, MA).

Brungart, D. (**1999**). ''Auditory localization of nearby sources. III. Stimulus effects,'' J. Acoust. Soc. Am. **106**, 3589–3602.

Brungart, D. S., and Rabinowitz, W. M. (**1999**). ''Auditory localization of nearby sources. Head-related transfer functions,'' J. Acoust. Soc. Am. **106**, 1465–1479.

Brungart, D. S., Durlach, N. I., and Rabinowitz, W. M. (**1999**). ''Auditory localization of nearby sources. II. Localization of a broadband source,'' J. Acoust. Soc. Am. **106**, 1956–1968.

Buell, T. N., Trahiotis, C., and Bernstein, L. R. (**1991**). ''Lateralization of low-frequency tones: Relative potency of gating and ongoing interaural delays,'' J. Acoust. Soc. Am. **90**, 3077–3085.

Butler, R. A., and Flannery, R. (**1980**). ''The spatial attributes of stimulus frequency and their role in monaural localization of sound in the horizontal plane,'' Percept. Psychophys. **28**, 449–457.

Carlile, S., Delaney, S., and Corderoy, A. (**1999**). ''The localisation of spectrally restricted sounds by human listeners,'' Hear. Res. **128**, 175–189.

Constan, Z. A., and Hartmann, W. M. (**2001**). ''Sound localization by interaural time differences at high frequencies,'' J. Acoust. Soc. Am. **109**, 2485.

Dau, T., Verhey, J., and Kohlrausch, A. (**1999**). ''Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers,'' J. Acoust. Soc. Am. **106**, 2752–2760.

Duda, R. O. (**1997**). ''Elevation dependence of the interaural transfer function,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. Gilkey and T. R. Anderson (Laurence Erlbaum, Mahwah, NJ), pp. 49–75.

Eberle, G., McAnally, K. I., Martin, R. L., and Flanagan, P. (**2000**). ''Localization of amplitude-modulated high-frequency noise,'' J. Acoust. Soc. Am. **107**, 3568–3571.

Fedderson, W. E., Sandel, T. T., Teas, D. C., and Jeffress, L. A. (**1957**). ''Localization of high-frequency tones,'' J. Acoust. Soc. Am. **29**, 988–991.

Freyman, R. L., Zurek, P. M., Balakrishnan, U., and Chiang, Y. C. (**1997**). ''Onset dominance in lateralization,'' J. Acoust. Soc. Am. **101**, 1649–1659.

Gaik, W. (**1993**). ''Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling,'' J. Acoust. Soc. Am. **94**, 98–110.

Hafter, E. R., and Carrier, S. C. (**1969**). ''Inability of listeners to trade completely interaural time for interaural intensity in a detection task,'' J. Acoust. Soc. Am. **49**, 125.

Hafter, E. R., and Dye, D. H. (**1983**). ''Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number,'' J. Acoust. Soc. Am. **73**, 644–651.

Hafter, E. R., and Jeffress, L. A. (**1968**). ''Two-image lateralization of tones and clicks,'' J. Acoust. Soc. Am. **44**, 563–569.

Hall, J. W., and Grose, J. H. (**1990**). ''Comodulation masking release and auditory grouping,'' J. Acoust. Soc. Am. **88**, 119–125.

Harris, G. G. (**1960**). ''Binaural interactions of impulsive stimuli and pure tones,'' J. Acoust. Soc. Am. **32**, 685–692.

Hartmann, W. M., and Constan, Z. A. (**1998**). ''Binaural coherence and the localization of sound in rooms,'' in *Proceedings of the 16th International Congress on Acoustics*, pp. 2889–2890.

Henning, G. B. (**1974**). ''Detectability of interaural delay in high-frequency complex waveforms,'' J. Acoust. Soc. Am. **55**, 84–90.

Hofman, P. M., and Van Opstal, A. J. (**1998**). ''Spectro-temporal factors in two-dimensional human sound localization,'' J. Acoust. Soc. Am. **103**, 2634–2648.

Humanski, R. A., and Butler, R. A. (**1988**). ''The contribution of the near and far ear towards localization of sound in the median plane,'' J. Acoust. Soc. Am. **83**, 2300–2310.

Kuhn, G. F. (**1987**). ''Physical acoustics and measurements pertaining to directional hearing,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 3–25.

Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (**1999**). ''Sensitivity of human subjects to head-related transfer-function phase spectra,'' J. Acoust. Soc. Am. **105**, 2821–2840.

Lawson, J. L., and Uhlenbeck, G. E. (**1950**). *Threshold Signals, Vol. 24 of Radiation Laboratory Series* (McGraw–Hill, New York).

Leakey, D. M., Sayers, B. M., and Cherry, C. (**1958**). ''Binaural fusion of low- and high-frequency sounds,'' J. Acoust. Soc. Am. **30**, 222.

Macpherson, E. A. (**1996**). ''Effects of source spectrum irregularity and uncertainty on sound localization,'' J. Acoust. Soc. Am. **99**, 2515–2529.

Macpherson, E. A. (**1998**). ''Spectral cue processing in the auditory localization of sounds with wideband non-flat spectra,'' Ph.D. thesis, University of Wisconsin—Madison.

Macpherson, E. A., and Middlebrooks, J. C. (**2000**). ''Localization of brief sounds: Effects of level and background noise,'' J. Acoust. Soc. Am. **108**, 1834–1849.

McFadden, D., and Pasanen, E. G. (**1976**). ''Lateralization at high frequencies based on interaural time differences,'' J. Acoust. Soc. Am. **59**, 634–639.

Middlebrooks, J. C. (**1999a**). ''Individual differences in external-ear transfer functions reduced by scaling in frequency,'' J. Acoust. Soc. Am. **106**, 1480–1492.

Middlebrooks, J. C. (**1999b**). ''Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency,'' J. Acoust. Soc. Am. **106**, 1493–1510.

Middlebrooks, J. C., and Green, D. M. (**1990**). ''Directional dependence of interaural envelope delays,'' J. Acoust. Soc. Am. **87**, 2149–2162.

Morimoto, M. (**2001**). ''The contribution of two ears to the perception of vertical angle in sagittal planes,'' J. Acoust. Soc. Am. **109**, 1596–1603.

Morimoto, M., and Aokata, H. (**1984**). ''Localization cues of sound sources in the upper hemisphere,'' J. Acoust. Soc. Jpn. **5**, 165–173.

Moushegian, G., and Jeffress, L. A. (**1959**). ''Role of interaural time and intensity differences in lateralization of low-frequency tones,'' J. Acoust. Soc. Am. **31**, 1441–1445.

Nuetzel, J. M., and Hafter, E. R. (**1981**). ''Discrimination of interaural delays in complex waveforms: Spectral effects,'' J. Acoust. Soc. Am. **69**, 1112–1118.

Palmer, A. R., and Russell, I. J. (**1986**). ''Phase-locking in the cochlear nerve of the guiniea-pig and its relation to the receptor potential of inner hair cells,'' Hear. Res. **24**, 1–15.

Pinheiro, M. L., and Tobin, H. (**1969**). ''Interaural intensity difference for intercranial lateralization,'' J. Acoust. Soc. Am. **46**, 1482–1487.

Rakerd, B., Hartmann, W. M., and McCaskey, T. L. (**1999**). ''Identification and localization of sound sources in the median saggital plane,'' J. Acoust. Soc. Am. **106**, 2812–2820.

Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). ''Phase locked responses to low-frequency tones in single auditory nerve fibers of the squirrel monkey,'' J. Neurophysiol. **30**, 769–793.

Saberi, K. (**1995**). ''Lateralization of comodulated complex waveforms,'' J. Acoust. Soc. Am. **98**, 3146–3156.

Sandel, T. T., Teas, D. C., Fedderson, W. E., and Jeffress, L. A. (**1955**). ''Localization of sound from single and paired sources,'' J. Acoust. Soc. Am. **27**, 842–852.

Searle, C. L., Braida, L. D., Cuddy, D. R., and Davis, M. F. (**1975**). ''Binaural pinna disparity: Another auditory localization cue,'' J. Acoust. Soc. Am. **57**, 448–455.

Simon, H. J., and Aleksandrovsky, I. (**1997**). ''Perceived lateral position of narrow-band noise in hearing-impaired and normal-hearing listeners under conditions of equal sensation level and sound-pressure level,'' J. Acoust. Soc. Am. **102**, 1821–1826.

Slaney, M. (**1994**). ''Auditory toolbox: A MATLAB toolbox for auditory modeling work,'' Technical Report 45, Apple Computer.

Slattery, W. H. I., and Middlebrooks, J. C. (**1994**). ''Monaural sound localization: Acute versus chronic impairment,'' Hear. Res. **75**, 38–46.
Strutt, J. W. (**1907**). ''On our perception of sound direction,'' Philos. Mag. **13**, 214–232.
Tobias, J. V., and Schubert, E. D. (**1959**). ''Effective onset duration of auditory stimuli,'' J. Acoust. Soc. Am. **31**, 1595–1605.
Trahiotis, C., and Bernstein, L. R. (**1986**). ''Lateralization of bands of noise and sinusoidally amplitude-modulated tones: Effects of spectral locus and bandwidth,'' J. Acoust. Soc. Am. **79**, 1950–1957.
Trahiotis, C., and Stern, R. M. (**1994**). ''Across-frequency interaction in lateralization of complex binaural stimuli,'' J. Acoust. Soc. Am. **96**, 3804–3806.
Whitworth, R. H., and Jeffress, L. A. (**1961**). ''Time vs intensity in the localization of tones,'' J. Acoust. Soc. Am. **33**, 925–929.
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. (**1997a**). ''Factors affecting the relative salience of sound localization cues,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Laurence Erlbaum, Mahwah, NJ), pp. 1–23.
Wightman, F. L., and Kistler, D. J. (**1997b**). ''Monaural localization revisited,'' J. Acoust. Soc. Am. **101**, 1050–1063.
Wightman, F. L., and Kistler, D. J. (**1999**). ''Sound localization with unilaterally degraded spectral cues,'' J. Acoust. Soc. Am. **105**, 1162.
Zhou, B., Green, D. M., and Middlebrooks, J. C. (**1992**). ''Characterization of external ear impulse responses using Golay codes,'' J. Acoust. Soc. Am. **92**, 1169–1171.
Zwislocki, J., and Feldman, R. S. (**1956**). ''Just noticeable differences in dichotic phase,'' J. Acoust. Soc. Am. **28**, 860–864.

Effect of delayed auditory feedback on normal speakers at two speech rates

Andrew Stuart,[a] Joseph Kalinowski, Michael P. Rastatter, and Kerry Lynch[b]
Department of Communication Sciences and Disorders, East Carolina University, Greenville, North Carolina 27858-4353

(Received 21 February 2001; accepted for publication 1 February 2002)

This study investigated the effect of short and long auditory feedback delays at two speech rates with normal speakers. Seventeen participants spoke under delayed auditory feedback (DAF) at 0, 25, 50, and 200 ms at normal and fast rates of speech. Significantly two to three times more dysfluencies were displayed at 200 ms ($p<0.05$) relative to no delay or the shorter delays. There were significantly more dysfluencies observed at the fast rate of speech ($p=0.028$). These findings implicate the peripheral feedback system(s) of fluent speakers for the disruptive effects of DAF on normal speech production at long auditory feedback delays. Considering the contrast in fluency/dysfluency exhibited between normal speakers and those who stutter at short and long delays, it appears that speech disruption of normal speakers under DAF is a poor analog of stuttering. [DOI: 10.1121/1.1466868]

PACS numbers: 43.70.Dn [AL]

I. INTRODUCTION

The powerful fluency-enhancing effects of delayed auditory feedback (DAF) among individuals who stutter are well documented (Bloodstein, 1995). In contrast, numerous experiments with normal speakers have shown that DAF produces disruptive effects on the speech. Such effects include speech errors (e.g., repetition of phonemes, syllables, or words), changes in speech rate/reading duration, prolonged voicing, increased vocal intensity, and modifications in aerodynamics (Black, 1951; Fukawa *et al.*, 1988; Howell, 1990; Langova *et al.*, 1970; Lee, 1950, 1951; Mackay, 1968; Siegel *et al.*, 1982; Stager *et al.*, 1997; Stager and Ludlow, 1993). Several theorists (Black, 1951; Cherry and Sayers, 1956; Van Riper, 1982; Yates, 1963) have proposed that the speech disruptions of normal speakers under DAF are an analog of stuttering since these disruptions are similar to stuttering. Put simply, normal speakers can be made to ''artificially stutter'' under DAF.

With respect to speech production errors, three problems become evident when reviewing previous research examining the effects of DAF on normal speakers. First, investigators have typically utilized ''long'' delays ranging 100 to 300 ms. While such long delays have induced errors in speech production in normal speakers, there is a paucity of information concerning the effect of shorter delays. Second, to the best of our knowledge, there is only one study investigating the effect of different rates of speaking (e.g., normal versus a fast rate) and DAF on normal speakers. Zanini *et al.* (1999) reported that participants speaking at a normal rate while receiving 200 ms DAF produced significantly more speech errors that those receiving no DAF. With an increased speaking rate, the total number of speech errors increased for those receiving no DAF but remained approximately the same for those receiving DAF. There was no significant difference in speech errors at an increased speaking rate between those receiving DAF and those not. There is no evidence of the effect of speech rate and DAF at shorter delays. Finally, the absence of an operational definition of ''errors in speech production'' or ''dysfluency'' makes interpretation of earlier work problematic and most likely impossible. Specifically, definitions for dysfluency such as ''misarticulations'' (Ham *et al.*, 1984), ''hesitations'' (Stephen and Haggard, 1980), or ''slurred syllables'' (Zalosh and Salzman, 1965) are not consistent with the standard definition of dysfluent behaviors of individuals who stutter (i.e., part word repetitions, prolongations, and postural fixations).

This investigation sought to further explore the effect of DAF on normal speakers. Specifically, the purpose of this study was to investigate the effect of short and long auditory feedback delays at fast and normal rates of speech with normal speakers. In contrast to previous research, a conventional definition of dysfluency, consistent with the operational construct used in the examination of the dysfluency in those that stutter, was adopted. This definition excluded speech errors that are associated with other pathological conditions (i.e., developmental articulation errors).

II. METHOD

A. Participants

Seventeen normal speaking adult males aged 19 to 57 ($M=32.9$ years, s.d.$=12.5$) served as participants. All participants presented with normal middle ear function (American Speech-Language-Hearing Association, 1997) and normal hearing sensitivity defined as having pure-tone thresholds at octave frequencies from 250 to 8000 Hz and speech recognition thresholds of $\leqslant$20 dB HL (ANSI, 1996). All individuals had a negative history of neurological, otological, and psychiatric disorders.

[a]Electronic mail: stuarta@mail.ecu.edu
[b]Currently affiliated with Powhatan County Public Schools, Powhatan, VA.

B. Apparatus and procedure

All testing was conducted in an audiometric test suite. Participants spoke into a microphone (Shure Prologue Model 12L-LC) which the output was fed to an audio mixer (Mackie Micro Series 1202) and routed to a digital signal processor (Yamaha Model DSP-1) and amplifier (Optimus Model STA-3180) before being returned bilaterally through earphones (EAR Tone Model 3A). The digital signal processor introduced feedback delays of 0, 25, 50, or 200 ms to the participants' speech signal. The shorter delays were identical to those utilized by Kalinowski *et al.* (1996) with persons who stutter. The 200-ms delay was chosen to be representative of a long delay that was employed in numerous previous studies with normal speakers. The output to the earphones was calibrated to approximate real ear average conversation sound pressure levels of speech outputs from normal-hearing participants. All speech samples were recorded with a video camera (JVC Model S-62U) and a stereo videocassette recorder (Samsung Model VR 8705).

Participants read passages of 300 syllables with similar theme and syntactic complexity. Passages were read at both normal and fast speech rates under each DAF condition. Participants were instructed to read with normal vocal intensity. For the fast rate condition, participants were instructed to read as fast as possible while maintaining intelligibility. Speech rates were counterbalanced and DAF conditions were randomized across participants.

The number of dysfluent episodes and speech rates were determined for each experimental condition by trained research assistants. A dysfluent episode was defined as a part-word prolongation, part-word repetition, or inaudible postural fixation (i.e., ''silent blocks;'' Stuart *et al.*, 1997). The same research assistant recalculated dysfluencies for 10% of the speech samples chosen at random. Intrajudge syllable-by-syllable agreement was 0.92, as indexed by Cohen's *kappa* (Cohen, 1960). Cohen's *kappa* values above 0.75 represent excellent agreement beyond chance (Fleiss, 1981). A second research assistant independently determined stuttering frequency for 10% of the speech samples chosen at random. Interjudge syllable-by-syllable agreement was 0.89 as indexed by Cohen's *kappa*. Speech rate was calculated by transferring portions of the audio track recordings onto a personal computer's (Apple Power Macintosh 9600/300) hard drive via the videocassette recorder interfaced with an analog to digital input/output board (Digidesign Model Audiomedia NuBus). Sampling frequency and quantization were 22 050 Hz and 16 bit, respectively. Speaking rate was determined from samples of 50 perceptually fluent syllables that were contiguous and separated from dysfluent episodes by at least one syllable. Sample duration represented the time between acoustic onset of the first syllable and the acoustic offset of the last fluent syllable, minus pauses that exceeded 0.1 s. Most pauses were inspiratory gestures with durations of approximately 0.3 to 0.8 s. Speech rate, in syllables/second, was calculated by dividing the number of syllables in the sample by the duration of each fluent speech sample.

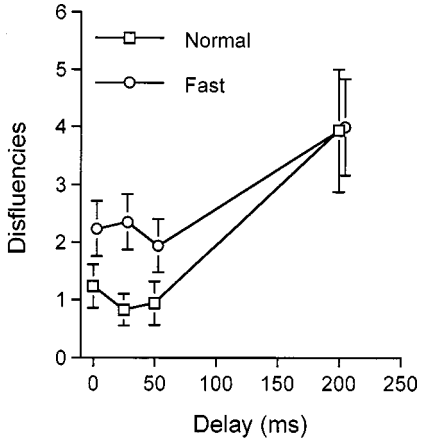


FIG. 1. Mean number of dysfluencies as a function of DAF and speech rate. Error bars represent plus/minus one standard error of the mean.

III. RESULTS

Means and standard deviations for dysfluencies (i.e., number of dysfluent episodes/300 syllables) as a function of DAF and speech rate are shown in Fig. 1. A two-factor analysis of variance with repeated measures was performed to investigate the effect of DAF and speech rate on dysfluencies. Statistically significant main effects of DAF [$F(3,48)=8.73$, Huynh-Felt $p=0.0015$, $\eta^2=0.35$] and speech rate [$F(1,16)=5.88$, Huynh-Felt $p=0.028$, $\eta^2=0.27$] were found. The effect sizes of these significant main effects were large (Cohen, 1988). The interaction of speech rate by DAF was not significant [$F(3,48)=1.10$, Huynh-Felt $p=0.33$, $\eta^2=0.064$, $\phi=0.20$ at $\alpha=0.05$]. *Post-hoc* orthogonal single-*df* contrasts showed that while the mean differences in dysfluencies at 0, 25, and 50 ms were not significantly different from each other ($p>0.05$) they were all significantly less than that at 200 ms ($p<0.05$).

Mean syllable rates and standard deviations as a function of DAF and speech rate are displayed in Fig. 2. A two-factor analysis of variance with repeated measures were performed to investigate the effect of DAF and speaking rate on syllable rate. Statistically significant main effects of DAF [$F(3,48)=39.32$, Huynh-Felt $p<0.0001$, $\eta^2=0.71$] and speaking rate condition [$F(1,16)=31.98$, Huynh-Felt p

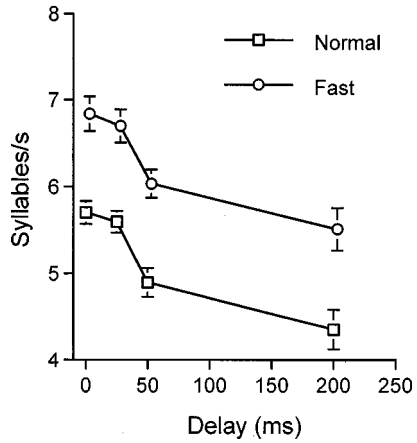


FIG. 2. Mean speech rates (in syllables/s) as a function of DAF and speech rate. Error bars represent plus/minus one standard error of the mean.

<0.0001, $\eta^2=0.66$] were found. The effect sizes of these significant main effects were large (Cohen, 1988). A nonsignificant DAF by speaking rate condition was found [$F(3,48)=0.02$, Huynh-Felt $p=0.99$, $\eta^2=0.001$, $\phi=0.054$ at $\alpha=0.05$]. *Post-hoc* orthogonal single-*df* comparisons revealed that there was no significant difference between syllable rates at 0 and 25 ms ($p>0.05$), they were significantly greater than 50 and 200 ms syllable rates, and the 50 ms was significantly greater than the 200 ms syllable rate ($p<.05$). In other words, participants were able to increase syllable rate when they were asked to speak fast under all DAF conditions. Participants decreased syllable rate at 50 and 200 ms during both speech rates relative to 0 and 25 ms DAF.

IV. DISCUSSION AND CONCLUSIONS

The present findings are threefold: First, DAF induced more significantly more dysfluencies only at the longest delay (i.e., 200 ms). In other words, normal speakers were capable of producing fluent or nearly fluent speech with short auditory feedback delays (i.e., $\leqslant$50 ms) that were equivalent to speech produced with no delay (i.e., 0 ms). Second, more dysfluencies were evident at a fast rate of speech. This finding would be consistent with increased motor load (Abbs and Cole, 1982; Borden, 1979; Borden and Harris, 1984). Finally, consistent with previous research (Black, 1951; Ham *et al.*, 1984; Lee, 1950; Siegel *et al.*, 1982; Stager and Ludlow, 1993), reduced speech rate was evidenced at auditory feedback delays greater than 25 ms with a greater reduction in syllable rate with an increase in DAF (i.e., 200 relative to 50 ms).

These findings suggest that temporal alterations in auditory feedback signal impact the speech-motor control system differentially for people who stutter and those that do not. That is, at delays of $\geqslant$50 ms individuals who stutter experience significant reductions (i.e., approximately 90%) in stuttering frequency (e.g., Kalinowski *et al.*, 1996) while in contrast normal speakers begin to experience dysfluent behavior at delays of $>$50 ms. What remains is a parsimonious explanation for two apparent paradoxical effects in altered auditory feedback.

Models of normal and stuttered speech production/monitoring have generally discounted the role of auditory feedback of having any significant role or any direct impact on central speech production commands since it is too slow (Borden, 1979; Levelt, 1983, 1989). As recognition of running speech is possible only at approximately 200 ms following production (Marslen-Wilson and Tyler, 1981, 1983), one could suggest that it should be of no surprise that the disruption of running speech production does not occur at auditory feedback of delays less than 200 ms in normal speakers. That is, peripheral feedback mechanisms (audition, taction, and/or proprioception) are affecting central speech motor control.

What then is the role of DAF in reducing dysfluency in those who stutter? It was generally posited that the stuttering reducing properties of DAF were due to an altered manner of speaking, specifically syllable prolongation, and not to any antecedent in the auditory system (Costello-Ingham, 1993; Perkins, 1979; Wingate, 1976). However, the role of the auditory system and DAF was revised by Kalinowski *et al.* (1996) who suggested that if a slow speech rate was necessary for stuttering reduction, then the stuttering reducing properties of DAF should not be evident when individuals who stutter speak at a fast speech rate. They had individuals who stutter read passages under conditions of altered auditory feedback including DAF at normal and fast rates of speech. Their results showed that stuttering episodes decreased significantly by approximately 70% under DAF regardless of speaking rate. These findings contradicted the notion regarding the importance of syllable prolongation to fluency induced by DAF. It was not suggested that syllable prolongation is unimportant to stuttering reduction *per se*, but rather, when syllable prolongation is eliminated, such as when speaking at a fast rate, the stuttering reduction properties of DAF are just as robust and can be most likely attributed to their impact on the auditory system.

How then can DAF impact the auditory system of individuals who stutter, particularly at short delays? Recent findings from brain imaging studies provide some answers. Magnetoencephalography (MEG) offers excellent temporal resolution (i.e., ms) in the analysis of cerebral processing in response to auditory stimulation. It has been known for more than a decade that a robust response (M100) is generated in the supratemporal auditory cortex in response to auditory stimuli beginning 20 to 30 ms and peaking approximately 100 ms after stimulus onset (Näätänen and Picton, 1987). More recently it has been demonstrated that an individual's own utterances can reduce the M100 response. Curio *et al.* (2000) examined such during a speech/replay task. In the speech condition participants uttered two vowels in a series while listening to a random series of two tones. In the replay condition the same participants listened to the recorded vowel utterances from the speech condition. The self-produced recorded vowels evoked the M100 response in the replay condition. More interestingly, this response was significantly delayed in both auditory cortices and reduced in amplitude prominently in the left auditory cortex during speech production of the same utterances in the speech condition. Similar findings of inhibition of cortical neurons have been found with primates during phonation (Müller-Preuss *et al.*, 1980; Müller-Preuss and Ploog, 1981). These data have been interpreted to indicate central motor-to-speech priming in the form of inhibition of the auditory cortices during speech production (Curio *et al.*, 2000).

The implications of these findings can lead one to speculate that this motor-to-speech priming may be defective in individuals who stutter. There is evidence to suggest that this is the case: Salmelin *et al.* (1998) reported in another MEG study that the functional organization of the auditory cortex is different in those who stutter relative to normal fluent speakers. MEG was recorded while individuals who stutter and matched controls read silently, read with oral movement but without sound, read aloud, and in chorus with another while listening to tones delivered alternately to the left and right ears. M100 responses were the same in the two silent conditions but delayed and reduced in amplitude during the two spoken conditions. Although the temporal response of the M100 was similar between the two groups, response amplitude was not. An unusual interhemispheric balance was

evident with the participants who stuttered. The authors reported ''rather paradoxically, dysfluency was most likely to occur when the hemispheric balance in stutterers become more like that in normal controls...dysfluent *versus* fluent reading conditions in stutterers were associated with differences specifically in the left auditory cortex...[and] source topography also differed in the left hemisphere'' (p. 2229). It has been suggested that suppression and/or delay of the M100 response during tasks reflects a diminution in the number or synchrony of auditory cortical neurons available for processing auditory input—in the case of speech production and perception (Hari, 1990; Näätänen and Picton, 1987). Salmelin *et al.* (1998) suggested that the interhemispheric balance is less stable in those who stutter and may be more easily unhinged with an increased work load (i.e., speech production). Disturbances may cause transient unpredictable disruptions in auditory perception (i.e., motor-to-speech priming after Curio *et al.*, 2000) that could initiate stuttering. Salmelin *et al.* (1998) pointedly remarked that, during choral reading where all participants who stutter were fluent, left hemispheric sensitivity was restored. This may be the case with all fluency-enhancing conditions of altered auditory feedback including DAF. The left auditory cortex as the locus of discrepancy between fluent speakers and those with stuttering has been implicated in numerous other brain imaging studies (e.g., Braun *et al.*, 1997; De Nil *et al.*, 2000; Fox *et al.*, 2000; Wu *et al.*, 1995). There is also recent converging evidence implicating anomalous anatomy (i.e., planum temporal and posterior superior temporal gyrus) in persons who stutter (Foundas *et al.*, 2001). It remains to be seen if this is a cause or effect of stuttering. Further research is warranted.

Finally, considering the contrast in fluency/dysfluency exhibited between normal speakers and those who stutter and the differences in the functional organization in the brain between individuals who stutter and fluent speakers, it appears that speech disruption of normal speakers under DAF is a poor analog of stuttering. MEG studies have implicated the role of the auditory system on a central level and on a time scale compatible with the behavioral effects of DAF on the overt manifestations of the disorder. The data herein implicate the peripheral feedback system(s) of fluent speakers for the disruptive effects of DAF on normal speech production.

ACKNOWLEDGMENT

This work was presented in part at the 1998 American Speech-Language-Hearing Association Annual Convention, San Antonio, 20 November 1998.

Abbs, J. H., and Cole, K. J. (**1982**). ''Consideration of bulbar and suprabulbar afferent influences upon speech motor coordination and programming,'' in Speech Motor Control, edited by S. Grillner, B. Lindblom, J. Lubker, and A. Persson (Pergamon, Oxford, UK), pp. 159–186.

ANSI (**1996**). ANSI S3.6-1996, ''Specifications for audiometers'' (American National Standards Institute, New York).

American-Speech-Language-Hearing Association (**1997**). Guidelines For Audiologic Screening (American-Speech-Language-Hearing Association, Rockville Pike, MD).

Black, J. (**1951**). ''The effect of side-tone delay upon vocal rate and intensity,'' J. Speech Hear Disord. **16**, 50–56.

Bloodstein, O. (**1995**). *A Handbook on Stuttering*, 5th ed. (National Easter Seal Society, Chicago), pp. 327–357.

Borden, G. J. (**1979**). ''An interpretation of research on feedback interruption in speech,'' Brain Lang. **7**, 307–319.

Borden, G. J., and Harris, K. S. (**1984**). *Speech Science Primer* (Williams and Wilkins, Baltimore, MD).

Braun, A. R., Varga, M., Stager, S., Schulz, G., Selbie, S., Maisog, J. M., Carson, R. E., and Ludlow, C. L. (**1997**). ''Altered patterns of cerebral activity during speech and language production in developmental stuttering. An $H_2{}^{(15)}O$ positron emission tomography study,'' Brain **120**, 761–784.

Cherry, E., and Sayers, B. (**1956**). ''Experiments upon total inhibition of stammering by external control and some clinical results,'' J. Psychosom. Res. **1**, 233–246.

Cohen, J. (**1960**). ''A coefficient of agreement for nominal scales,'' Educ. Psychol. Meas. **20**, 37–46.

Cohen, J. (**1988**). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Erlbaum, Hillsdale, NJ), pp. 273–406.

Costello-Ingham, J. C. (**1993**). ''Current status of stuttering and behavior modification-I: Recent trends in the application of behavior modification in children and adults,'' J. Fluency Disord. **18**, 27–55.

Curio, G., Neuloh, G., Numminen, J., Jousmaki, V., and Hari, R. (**2000**). ''Speaking modifies voice-evoked activity in the human auditory cortex,'' Hum. Brain Mapp. **9**, 183–191.

De Nil, L. F., Kroll, R. M., Kapur, S., and Houle, S. (**2000**). ''A positron emission tomography study of silent and oral single word reading in stuttering and nonstuttering adults,'' J. Speech Hear. Res. **43**, 1038–1053.

Fleiss, J. L. (**1981**). *Statistical Methods for Rates and Proportions*, 2nd ed. (Wiley, New York), pp. 212–236.

Foundas, A. L., Bollich, A. M., Corey, D. M., Hurley, M., and Heilman, K. M. (**2001**). ''Anomalous anatomy of speech-language areas in adults with persistent developmental stuttering,'' Neurology **57**, 207–215.

Fox, P. T., Ingham, R. J., Ingham, J. C., Zamarripa, F., Xiong, J. H., and Lancaster, J. L. (**2000**). ''Brain correlates of stuttering and syllable production. A PET performance-correlation analysis,'' Brain **123**, 1985–2004.

Fukawa, T., Yoshioka, H., Ozawa, E., and Yoshida, S. (**1988**). ''Difference of susceptibility to delayed auditory feedback between stutterers and nonstutterers,'' J. Speech Hear. Res. **31**, 475–479.

Ham, R., Fucci, D., Cantrell, J., and Harris, D. (**1984**). ''Residual effect of delayed auditory feedback on normal speaking rate and fluency,'' Percept. Mot. Skills **59**, 61–62.

Hari, R. (**1990**). ''The neuromagnetic method in the study of the human auditory cortex,'' Adv. Audiol. **6**, 222–282.

Howell, P. (**1990**). ''Changes in voice level caused by several forms of altered auditory feedback in fluent speakers and stutterers,'' Lang Speech **33**, 325–338.

Kalinowski, J., Stuart, A., Sark, S., and Armson, J. (**1996**). ''Stuttering amelioration at various feedback delays and speech rates,'' Eur. J. Disord. Commun. **31**, 259–269.

Langova, J., Moravek, M., Novak, A., and Petrik, M. (**1970**). ''Experimental interference with auditory feedback,'' Folia Phoniatr. **22**, 191–196.

Lee, B. S. (**1950**). ''Effects of delayed speech feedback,'' J. Acoust. Soc. Am. **22**, 824–826.

Lee, B. S. (**1951**). ''Artificial stutterer,'' J. Speech Hear. Disord. **16**, 63–65.

Levelt, W. J. M. (**1983**). ''Monitoring and self-repair in speech,'' Cognition **14**, 41–104.

Levelt, W. J. M. (**1989**). *Speaking: From Intention to Articulation* (MIT, Cambridge, MA).

Mackay, D. (**1968**). ''Metamorphosis of critical interval: Age linked changes in the delay in auditory feedback that produces maximal disruptions of speech,'' J. Acoust. Soc. Am. **43**, 811–821.

Marslen-Wilson, W. D., and Tyler, L. K. (**1981**). ''Central process in speech understanding,'' Philos. Trans. R. Soc. London, Ser. B **259**, 297–313.

Marslen-Wilson, W. D., and Tyler, L. K. (**1983**). ''Reply to Cowart,'' Cognition **15**, 227–235.

Müller-Preuss, P., and Ploog, D. (**1981**). ''Inhibition of auditory cortical neurons during phonation,'' Brain Res. **215**, 61–76.

Müller-Preuss, P., Newman, J. D., and Jürgens, U. (**1980**). ''Anatomical and physiological evidence for a relationship between the 'cingular' vocalization area and the auditory cortex in the squirrel monkey,'' Brain Res. **202**, 307–315.

Näätänen, R., and Picton, T. (**1987**). ''The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure,'' Psychophysiology **24**, 375–425.

Perkins, W. H. (**1979**). ''From psychoanalysis to discoordination,'' in *Controversies About Stuttering Therapy*, edited by H. H. Gregory (University Park, Baltimore), pp. 97–127.
Salmelin, R., Schnitzler, A., Schmitz, F., Jäncke, L., Witte, O. W., and Freund, H.-J. (**1998**). ''Functional organization of the auditory cortex is different in stutterers and fluent speakers,'' NeuroReport **9**, 2225–2229.
Siegel, G. M., Schork, E. J., Jr., Pick, H. L., Jr., and Garber, S. R. (**1982**). ''Parameters of auditory feedback,'' J. Speech Hear. Res. **25**, 473–475.
Stager, S. V., and Ludlow, C. L. (**1993**). ''Speech production changes under fluency-evoking conditions in nonstuttering speakers,'' J. Speech Hear. Res. **36**, 245–253.
Stager, S. V., Denman, D. W., and Ludlow, C. L. (**1997**). ''Modifications in aerodynamic variables by persons who stutter under fluency-evoking conditions,'' J. Speech Hear. Res. **40**, 832–847.
Stephen, S. C. G., and Haggard, M. P. (**1980**). ''Acoustic properties of masking/delayed feedback in the fluency of stutterers and controls,'' J. Speech Hear. Res. **23**, 527–538.
Stuart, A., Kalinowski, J., and Rastatter, M. P. (**1997**). ''Effects of monaural and binaural altered auditory feedback on stuttering frequency,'' J. Acoust. Soc. Am. **101**, 3806–3809.
Van Riper, C. (**1982**). *The Nature of Stuttering*, 2nd ed. (Prentice Hall, Englewood Cliffs, NJ).
Wingate, M. E. (**1976**). *Stuttering: Theory and Treatment* (Irvington, New York).
Wu, J. C., Maguire, G., Riley, G., Fallon, J., LaCasse, L., Chin, S., Klein, E., Tang, C., Cadwell, S., and Lottenberg, S. (**1995**). ''A positron emission tomography [^{18}F] deoxyglucose study of developmental stuttering,'' NeuroReport **6**, 501–505.
Yates, A. J. (**1963**). ''Delayed auditory feedback,'' Psychol. Bull. **60**, 213–232.
Zalosh, S., and Salzman, L. F. (**1965**). ''After effects of delayed auditory feedback,'' Percept. Mot. Skills **20**, 817–823.
Zanini, S., Clarici, A., Fabbro, F., and Bava, A. (**1999**). ''Speaking speed effects on delayed auditory feedback disruption of speech fluency,'' Percept. Mot. Skills **89**, 1095–1109.

Children's use of semantic cues in degraded listening environments

Marianne Fallon, Sandra E. Trehub,[a] and Bruce A. Schneider
Department of Psychology, University of Toronto at Mississauga, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada

(Received 22 March 2001; accepted for publication 6 February 2002)

Children 5 and 9 years of age and adults were required to identify the final words of low- and high-context sentences in background noise. Age-related differences in the audibility of speech signals were minimized by selecting signal-to-noise ratios (SNRs) that yielded 78% correct performance for low-context sentences. As expected, children required more favorable SNRs than adults to achieve comparable levels of performance. A more difficult listening condition was generated by adding 2 dB of noise. In general, 5-year-olds performed more poorly than did 9-year-olds and adults. Listeners of all ages, however, showed comparable gains from context in both levels of noise, indicating that noise does not impede children's use of contextual cues. [DOI: 10.1121/1.1466873]

PACS numbers: 43.71.Ft, 43.71.Pc [CWT]

I. INTRODUCTION

When identifying words in spoken utterances, listeners typically monitor contextual cues as well as acoustic-phonetic information. At times, acoustic-phonetic cues are insufficient for the identification of particular words because of limitations in the signal (e.g., poor articulation on the part of the speaker) or listening environment (e.g., the presence of competing sounds or reverberation). In such circumstances, contextual cues—lexical, semantic, and syntactic—can help listeners decode a message. For example, upon hearing the word ''eating,'' listeners expect a food-related noun to follow. Although adults derive considerable benefit from such contextual cues, especially in challenging listening environments (e.g., Kalikow *et al.*, 1977; Bilger *et al.*, 1984; Pichora-Fuller *et al.*, 1995; Dubno *et al.*, 2000), the extent to which children do so is unclear.

A number of investigators have examined children's use of semantic contextual cues to decode a degraded signal. For example, Cole and Perfetti (1980) presented mispronounced words in high-predictability and low-predictability sentences within a story. They found that 4-year-olds were better at detecting mispronounced words in high-predictability than in low-predictability sentences, indicating that very young children profit from such contextual cues. Craig *et al.* (1993) presented their target words in high-predictability contexts (e.g., ''The watchdog gave a warning growl'') or in low-predictability contexts (e.g., ''I had not thought about the growl''). They altered the to-be-identified word by eliminating the later portions of the word. In such ''gating'' tasks (Grosjean, 1980, 1985), participants engage in repeated attempts to identify the target word from successively longer portions, or gates. Craig *et al.* (1993) found that 8- to 10-year-old children required shorter gates (i.e., less acoustic information) to identify target words in high-predictability than in low-predictability sentences. By contrast, 5- to 7-year-olds required as much acoustic-phonetic information with high-predictability as with low-predictability contexts (Craig *et al.*, 1993). Because of young children's limitations in phonetic representation (Walley, 1988, 1993) and metalinguistic knowledge (Liu *et al.*, 1997), the gating task may underestimate their ability to use semantic contextual cues.

Tyler and Marslen-Wilson (1981) examined children's use of semantic contextual cues by altering the semantic adequacy of sentences. They presented target words in semantically appropriate (''John has to go back home'') or in anomalous (''John had to sit on the shop'') utterances. Children 5–10 years of age more readily detected the target words in semantically appropriate than in anomalous sentences, confirming their ability to profit from the semantic context. Liu *et al.* (1997) found that preschoolers repeat final target words more rapidly when they are presented in semantically appropriate contexts than in anomalous sentences. Even with compressed or filtered speech, 4-year-olds respond more quickly to semantically appropriate than to inappropriate sentences (Bates *et al.*, 2001).

There are recent indications that the ability to use semantic contextual cues emerges by 2 years of age. Fernald (2001) examined 2-year-olds' eye movements to two pictures as they listened to simple sentences with verbs that cued the target words (e.g., ''Drink the milk''). Eye movements toward the correct picture (i.e., the one cued by the semantic context) were initiated before the target word was presented. It is clear, then, that listening to speech engages predictive processing at relatively early stages of language development.

Despite evidence that young children benefit from semantic contextual cues in optimal listening conditions (e.g., Tyler and Marslen-Wilson, 1981; Fernald, 2001) or with compressed or filtered speech (e.g., Bates *et al.*, 2001), there are suggestions that background noise interferes with young children's use of such cues. Nittrouer and Boothroyd (1990) presented short sentences varying in semantic adequacy in a background of spectrally matched noise. The listener's task was to repeat the sentences verbatim. Although semantic cues facilitated performance, 4- to 6-year-old children did not benefit from such cues to the same extent as did young and elderly adults. Nittrouer and Boothroyd (1990) speculate

[a]Electronic mail: sandra.trehub@utoronto.ca

that young children's limited knowledge of semantic constraints rather than the presence of noise was responsible for their lesser benefit from semantic cues. Elliott (1979) investigated children's use of semantic cues by means of the SPIN (speech perception in noise) task (Kalikow *et al.*, 1977; Bilger *et al.*, 1984). Participants in the SPIN task are required to repeat the final word of high- and low-predictability sentences presented in multitalker babble. Elliott (1979) found that children 9–13 years of age derived less benefit from contextual cues in high-predictability sentences than did older listeners. Interestingly, the ability to profit from semantic cues (as assessed by the SPIN task) continues to improve into older adulthood (Pichora-Fuller *et al.*, 1995; but see Dubno *et al.*, 2000). Pichora-Fuller *et al.* (1995) suggest that elderly adults may compensate for a declining perceptual system by relying on contextual cues to a greater extent than younger listeners do.

In principle, children whose perceptual system is not fully developed have much to gain from semantic contextual cues, especially in noisy backgrounds. The available evidence (e.g., Elliott, 1979; Nittrouer and Boothroyd, 1990) indicates, however, that they do not profit optimally from such cues. Why not? A number of factors may be implicated. Young children are likely to perform more poorly than adults on difficult listening tasks simply on the basis of their 5-dB difference in absolute and masked thresholds (e.g., Schneider *et al.*, 1986, 1989). For example, Fallon *et al.* (2000) found that 5-year-old children required signal-to-noise ratios (SNRs) that were 5 dB more favorable than those of adults to achieve comparable identification in simple low-context sentences (e.g., ''Touch the dog''). Failure to compensate for age-related sensory differences obscures the independent contributions of sensory and cognitive factors. Second, the test materials in Nittrouer and Boothroyd (1990) and Elliott (1979) were not designed for young children. As a result, some of Nittrouer and Boothroyd's (1990) high-predictability sentences (e.g., ''Tough guys sound mean;'' ''Dull paint won't shine'') may have confused many, if not all, young children. Third, the articulatory demands of repeating the final word of lengthy test sentences (Elliott, 1979) may have posed disproportionate difficulty for very young children (Elliott *et al.*, 1979). In short, meaningful comparisons of children's and adults' use of contextual cues in noise depend on comparable task demands and listening conditions across age.

In the present study, we evaluated the contribution of semantic contextual cues to word recognition in noise by 5- and 9-year-old children and adults. To this end, we devised age-appropriate high- and low-context sentences in which the target word was a familiar monosyllabic noun. Task difficulty was minimized by means of a four-alternative picture-pointing response (following Fallon *et al.*, 2000). To minimize age-related differences in stimulus audibility, we established SNRs that yielded comparable performance levels (78%) on low-context sentences (e.g., ''We looked at the bread''). Greater use of contextual cues by older children and adults than by younger children would yield a statistical interaction between age and context. A secondary goal was to evaluate the prevailing view that noise has more adverse perceptual consequences on young children than it does on older listeners (Mills, 1975; Elliott, 1979). Although Fallon *et al.* (2000) found that noise increments have comparable effects on 5-year-olds, 9-year-olds, and adults, their test sentences (e.g., ''Touch the X'') were structurally simpler than those in the current study. It is possible, then, that noise increments would reveal age differences in the processing of more complex sentences. Accordingly, separate groups of listeners were tested with different noise levels. Disproportionate consequences of noise increments on young listeners would result in an interaction between age and noise level.

We recorded latencies for correctly identified low- and high-context targets to complement the accuracy measures. Recall that semantically appropriate sentences promote faster word repetition than do semantically impoverished or anomalous sentences (Liu *et al.*, 1997). By extension, we expected faster word identification (i.e., pointing to the target picture) in high- than in low-context sentences. Such reductions in latency would indicate that listeners use semantic cues to predict forthcoming words. If 5-year-olds and adults use semantic context similarly, they should show comparable reductions in response latency to high-context sentences.

II. METHOD

A. Participants

The participants were 48 5-year-olds (range: 5;0–5;6, $M = 5;3$ years), 48 9-year-olds (range: 9;0–9;6, $M = 9;3$ years), and 48 adults (range: 19–25, $M = 20.7$ years), with equal numbers of males and females in each age group. No child in the sample had experienced frequent ear infections or had ever required pressure-equalizing tubes. All children and adults were free of colds on the day of testing. Participants were monolingual speakers of English, with the exception of three adults, who had learned English by 6 years of age and attended elementary school in North America. Additional participants were excluded because of experimenter error (one 5-year-old, two 9-year-olds, one adult), failure to complete the test session (one 5-year-old), failure to meet the training criterion (four 5-year-olds), failure to follow instructions (one 5-year-old, one 9-year-old), inattentiveness (three 5-year-olds, one 9-year-old), or apprehensiveness (two 5-year-olds).

B. Apparatus and stimuli

Testing occurred in a double-wall sound-attenuating booth, 3 m×2.8 m×2 m in size. Participants were seated facing a nonglare touch-screen monitor (Goldstar 1465DLs) 33×33 cm. Loudspeakers (KEF Model 101) were 45 degrees to the left and right of the participant (distance of 70 cm) at approximate ear level. Sentence files and babble files were converted to analog form by means of Tucker-Davis digital-to-analog converters under the control of a computer with a Pentium processor. Sentence and babble amplitudes were controlled separately by means of programmable attenuators. After mixing, the combined signals were amplified (SAE 2600) and presented over loudspeakers located inside the testing booth. Sound-field levels were determined in the absence of the listener by means of a Bruel and Kjaer $\frac{1}{2}$-in. microphone.

TABLE I. Fundamental frequency and duration of low- and high-context sentences.

	Frequency (Hz)		Duration (ms)	
	Sentence	Target	Sentence	Target
Low context				
Mean	231.24	233.28	1787.43	605.26
SD	13.39	15.71	155.94	87.84
Range	198.30–257.65	200.15–262.27	1453.60–2040.20	432.15–783.25
High context				
Mean	231.12	238.02	2147.68	556.62
SD	9.36	20.55	354.04	79.57
Range	206.92–251.59	180.88–286.69	1500.15–2883.60	414.60–726.30

We generated a large corpus ($n=188$) of high-context sentences with concrete, monosyllabic nouns in sentence-final position. Sentences were relatively simple and short (5 to 10 syllables in length), with content that was familiar to 5-year-olds. To assess the degree to which high-context stems primed the final target words, an independent sample of 361 adults (289 women, 72 men; mean age=22.3 years; range 18–54) completed subsets of high-context sentence stems (e.g., ''Mice like to eat ___ '') with the word that first came to mind. Each participant completed approximately 47 sentence stems by means of a paper-and-pencil task, yielding approximately 90 observations per sentence. The data from 19 additional adults were excluded for failing to complete over 90% of the stems (four women, one man), learning English after age 14 (ten women, two men), or providing blatantly foolish answers (two men). From the initial corpus, 60 sentences were selected to serve as practice ($n=20$) and test ($n=40$) items. The mean number of syllables in high-context sentences was 7.28. On average, adults completed each high-context stem with the target word 52.5% of the time (range 3%–95%). Low-context sentences ($n=60$, 40 test items, 20 practice items) were generated with these target words. Low-context sentences could take one of five forms: (1) ''[Pronoun] looked at the [target]''; (2) ''[Pronoun] talked about the [target]''; (3) ''[Pronoun] read about the [target]''; (4) ''[Pronoun] heard about the [target]''; and (5) ''[Pronoun] pointed at the [target].'' Thus, low-context sentences contained 5 or 6 syllables ($M=5.77$). High- and low-context sentences are listed in the Appendix.

We verified 5-year-olds' ability to predict the target words from the high-context stems by means of a four-alternative, forced-choice task. Subsets ($n=20$) of prerecorded high-context sentence stems were presented at 44 dB (A) in quiet to an additional sample of 29 5-year-olds (14 girls, 15 boys; $M=5;3$ years). Sentence stems were accompanied by four pictures that appeared in different corners of a touch screen. The four alternatives included the target and three foils depicting words that were semantically and phonologically dissimilar from the target word. Visual feedback in the form of a flashing picture was provided for correct responses. The experimenter explained to children that they would hear a lady talking, but that the computer would ''chop off'' her last word. Their task was to select the picture corresponding to what the lady wanted to say. To ensure that children understood the instructions, they received two practice trials with photocopied arrays of pictures from practice plates of the *Peabody Picture Vocabulary Test* (Dunn and Dunn, 1997). For example, the experimenter presented pictures of a table, doll, car, and man and said, ''We put dinner on the ___.'' If children responded incorrectly on the practice trials, the experimenter provided the correct answer. Children selected the target word on 98.8% of test trials, indicating that the high-context sentence stems effectively cued the target words with ideal listening conditions and a closed-response set.

High- and low-context sentences were produced by a vocally trained woman and digitized at a rate of 20 kHz using a 16-bit Tucker Davis (DD1) analog-to-digital converter. The babble portion of SPIN forms used by Pichora-Fuller *et al.* (1995) was similarly digitized and stored. The sentences used in the main experiment (i.e., with target word included) had an average fundamental frequency of 231 Hz and an average duration of 1968 ms. As shown in Table I, the duration of high-context sentences exceeded that of low-context sentences, $t(78)=-5.56$, $p<0.001$, but the duration of target words in low-context sentences exceeded the duration of target words in high-context sentences, $t(78)=2.60$, $p=0.01$. The longer durations of low-context target words may reflect the speaker's intuitive compensation for low redundancy (Lieberman, 1963; Fowler, 1988). The mean fundamental frequency of target words did not differ as a function of contextual condition, $t(78)=-1.159$, $p=0.25$. Similarly, the mean fundamental frequency of low- and high-context sentences did not differ significantly, $t(78)=0.05$, $p=0.96$. Sentences were presented at approximately 44 dB(A), a level well above 5-year-olds' speech-identification threshold in forced-choice contexts. Pilot testing for a previous study (Fallon *et al.*, 2000) revealed that 5-year-olds identify monosyllabic target words in simple carrier phrases (e.g., ''Touch the X.'') with 99% accuracy. Root-mean-square (rms) values were calculated and adjusted such that each sentence was presented at equal intensity following the procedure of Schneider *et al.* (2000). SNR was varied by adjusting the level of babble (F_0: 185 Hz). Pilot-testing established the SNR at which each age group achieved approximately 78% correct performance on low-context sentences: −24 dB for 5-year-olds, −27 dB for 9-year-olds, and −30 dB for adults.[1] These levels were designated low-noise conditions. High-noise conditions were created for each age group by decreasing the SNR by 2 dB (by adding 2 dB of noise), resulting in SNRs of −26 dB for 5-year-olds, −29 dB for 9-year-olds, and −32 dB for adults. SNRs for the training

phase were set 5 dB higher than those in the low-noise conditions: −19, −22, and −25 dB for 5-year-olds, 9-year-olds, and adults, respectively.

Visual stimuli consisted of 60 black-and-white line drawings of familiar, concrete objects corresponding to the target items. Images were gathered from various sources, including Snodgrass and Vanderwart (1980), Cycowicz *et al.* (1997), the *Peabody Picture Vocabulary Test* (Dunn and Dunn, 1997), and a local artist.

C. Procedure

All participants were tested individually. A trial, which was initiated by a button press, consisted of the simultaneous presentation of vocal stimuli (sentence and noise) and visual stimuli. Sentences in noise were accompanied by an array of four different images, one appearing in each corner of the touch screen. The multitalker babble began with the onset of the sentence and terminated when the sentence ended. The visual array included the target image and three designated foils that were phonetically and semantically dissimilar from the target. The locations of targets and foils were randomly selected on each trial. Reaction times accurate to the millisecond were recorded automatically from stimulus offset. Feedback for correct performance consisted of the target flashing in the middle of the screen. Incorrect selections resulted in the screen going blank. Unlike the usual situation in reaction-time experiments, listeners were not told to respond as quickly as possible.

The instructions were tailored to the age of participants. Listeners were told that the last word of the sentence would correspond to one of the pictures on the screen. The experimenter explained to 5-year-old children that they should choose the picture that matches the last word that the lady says. Moreover, 5-year-olds were told that if they only hear part of a word, they should choose the picture that sounds similar to what they hear (e.g., ''If you hear 'irt' and there are pictures of a can, plate, shirt, and boat on the screen, you should pick the shirt because 'shirt' sounds the most like 'irt'.''). Older children and adults were told that the pictures were identifiable by the most basic term. For example, a picture of a shirt would be identified by the word ''shirt,'' not ''button-down'' or ''clothing.'' No other strategies were suggested.

The test session consisted of a training phase and a test phase. All participants were required to meet a training criterion of 4 consecutive correct responses within a maximum of 20 trials consisting of randomly ordered high- and low-context sentences. On average, listeners achieved the training criterion in 7.4 trials. After reaching the criterion, participants proceeded to the test phase, which consisted of 40 randomly ordered sentences (20 low-context, 20 high-context) presented in either low or high noise. Target items could occur only once within a test session, necessitating two lists (A and B), each containing 40 sentences. High- and low-context sentences were presented at identical SNRs. Half of the participants were tested in low noise, the other half in high noise. In each noise condition, half of the listeners received list A, with the remaining listeners receiving list B. Equal numbers of males and females in each age group were tested in each noise condition, and equal numbers received list A or B.

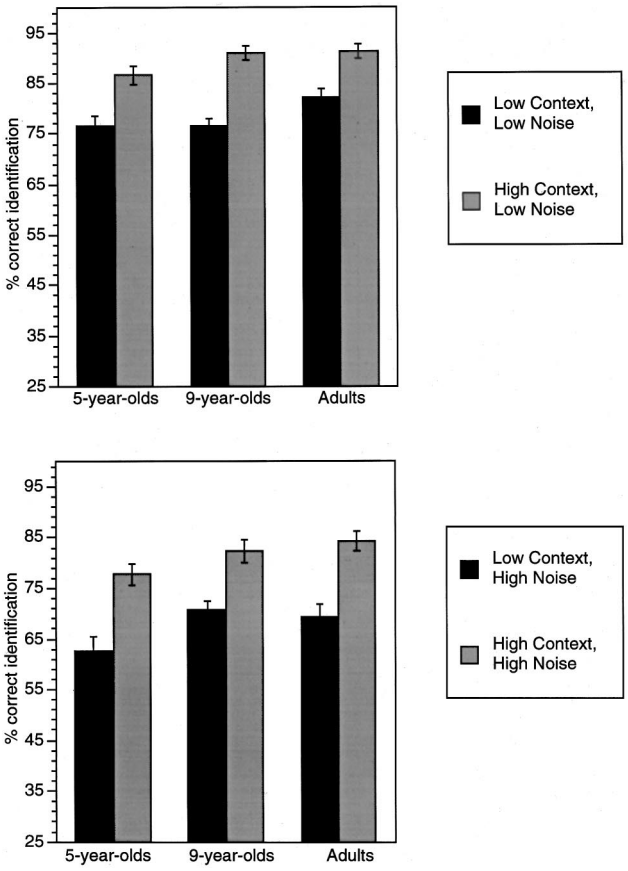


FIG. 1. Performance of 5-year-olds, 9-year-olds, and adults on low- and high-context sentences in both levels of noise. Error bars represent standard error of the mean.

Adults and 9-year-old children initiated trials at their preferred pace. The experimenter initiated trials for 5-year-olds when she judged them to be ready and attentive. The experimenter remained in the test booth during the entire session for children, offering periodic verbal reinforcement and encouragement that was unrelated to their performance. To maintain the interest of 5-year-olds, they received a colored sticker after every four trials, which they placed in an ''incomplete'' black-and-white picture. At the end of 40 trials, children had completed the picture.

III. RESULTS

Figure 1 illustrates performance at each noise level as a function of context and age. Because preliminary analyses revealed no difference in performance as a function of list (A or B), this factor was omitted from subsequent analyses. Despite our attempts to control for differences in stimulus audibility on low-context sentences, a one-way ANOVA with age as the independent variable revealed a marginal effect of age in the low-noise condition, $F(2,69)=2.78$, $p=0.07$. Post-hoc Tukey analyses, however, did not achieve conventional levels of significance ($p>0.1$ for all comparisons). A 2×3 ANOVA with context as a within-subject factor and age as a between-subject factor revealed that high-context sentences were identified more accurately than low-context sen-

TABLE II. Performance (percent correct) of 5-year-olds, 9-year-olds, and adults on low- and high-context sentences in both levels of noise.

	Low noise		High noise	
	Low context	High context	Low context	High context
5-year-olds	76.59 (9.51)	86.67 (9.17)	62.71 (13.90)	77.69 (10.30)
9-year-olds	76.46 (9.61)	91.04 (6.59)	70.83 (8.56)	82.29 (11.23)
Adults	82.08 (9.20)	91.25 (7.26)	69.38 (13.05)	84.17 (9.74)

Note: Standard deviations are in parentheses.

tences, $F(1,69)=79.94$, $p<0.001$, and that 5-year-olds performed more poorly than 9-year-olds and adults, $F(2,69)=3.34$, $p=0.041$ (see Table II). As can be seen in Fig. 1, there was no interaction between age and context, $F(2,69)=1.76$, $p=0.18$. Note, however, that performance on high-context sentences was positively skewed, with several participants (four 5-year-olds, three 9-year-olds, and four adults) achieving perfect scores.

In high noise, performance on low-context sentences varied significantly with age, $F(2,69)=3.09$, $p=0.052$. Post-hoc Tukey comparisons of 5- and 9-year-olds approached conventional levels of significance, $p=0.058$, as did those between 5-year-olds and adults, $p=0.087$ (see Table II). A 2×3 ANOVA with context as a within-subject factor and age as a between-subject factor revealed main effects of context, $F(1,69)=81.36$, $p<0.001$, and of age, $F(2,69)=3.92$, $p=0.024$, but no interaction between context and age, $F(2,69)=0.56$, $p=0.57$. Performance on high-context sentences in high noise was more normally distributed than was performance in low noise, minimizing the likelihood that interactions between age and context were obscured by ceiling effects.

To compare performance at the two noise levels, a $2\times2\times3$ repeated-measures ANOVA, with context as a within-subject factor and noise level and age as between-subject factors, revealed a main effect of context, $F(1,138)=160.01$, $p<0.0001$, and of age, $F(2,138)=6.59$, $p=0.002$, reflecting poorer performance by 5-year-olds ($M=75.83\%$, s.d.$=10.83$) than by adults ($M=81.72\%$, s.d.$=9.14$), Tukey HSD, $p=0.012$. As expected, listeners performed better in low noise ($M=83.99\%$, s.d.$=7.04$) than in high noise ($M=74.48\%$, s.d.$=9.62$), $F(1,138)=49.77$, $p<0.001$. There were no two-way or higher-order interactions. The same pattern of findings emerged from an analysis using rationalized arcsine transformations of the percent correct data (Studebaker, 1985), which indicates that ceiling effects were not responsible for obscuring an interaction between age and context.

To gain further perspective on context effects, we examined reaction times to words identified correctly in low- and high-context sentences across age. The means and standard deviations of individual listeners on low- and high-context sentences were calculated separately, excluding trials on which reaction times exceeded the individual's mean by 2 standard deviations. Discarded trials for low-context sentences represented 6.26% (s.d.$=3.87$), 5.32% (s.d.$=3.77$), and 5.23% (s.d.$=3.12$) of all correct trials for 5-year-olds, 9-year-olds, and adults, respectively. For high-context sentences, discarded trials represented 6.70% (s.d.$=3.34$), 6.33% (s.d.$=4.58$), and 5.88% (s.d.$=5.20$) of all correct trials for 5-year-olds, 9-year-olds, and adults, respectively. A repeated-measures ANOVA indicated that these discard rates did not differ across context, $F(1,138)=2.24$, $p>0.1$; age, $F(2,138)=1.28$, $p>0.2$; or noise, $F(1,138)=1.06$, $p>0.3$. No interactions between the factors reached conventional levels of significance. To achieve homogeneity of variance, two 5-year-old girls who exhibited very high latencies were excluded from the subsequent analysis. Table III reports average latencies in each age group as a function of context. A $2\times2\times2\times3$ repeated-measures ANOVA, with mean latency as the dependent variable, context as a within-subject factor, and age and noise as between-subject factors, revealed that word identification was faster in high-context sentences ($M=1.361$ s, s.d.$=0.573$) than in low-context sentences ($M=1.695$ s, s.d.$=0.573$), $F(1,130)=89.40$, $p<0.0001$. Moreover, word identification was faster in low noise ($M=1.449$ s, s.d.$=0.516$) than in high noise ($M=1.670$ s, s.d.$=0.544$), $F(1,130)=4.132$, $p=0.04$. As shown in Table III, 9-year-olds and adults responded more rapidly than 5-year-olds, $F(2,130)=7.84$, $p=0.001$, but there was no age $\times$context interaction.

TABLE III. Average latencies in seconds for low- and high-context sentences as a function of age.

	Low context	High context
5-year-olds	1.883 (0.595)	1.623 (0.625)
9-year-olds	1.672 (0.639)	1.254 (0.512)
Adults	1.534 (0.420)	1.217 (0.500)

Note: Standard deviations are in parentheses.

IV. DISCUSSION

Children 5 and 9 years of age and adults listened to high- and low-context sentences in noise (multitalker babble) at levels that were adjusted to equalize signal audibility across age. Although 5-year-olds experienced more difficulty than would be expected on the basis of age differences in sensitivity, listeners of all ages identified the target words more accurately and more quickly in high-context sentences than in low-context sentences and at lower noise levels than at higher noise levels. Moreover, gains in identification accuracy and reductions in latency were comparable in magnitude for all age groups.

Favorable effects of context on auditory word recognition are consistent with previous research on speech identification in noise by children (Elliott, 1979; Nittrouer and Boothroyd, 1990) and adults (e.g., Kalikow *et al.*, 1977; Bilger *et al.*, 1984; Boothroyd and Nittrouer, 1988). Nevertheless, comparable gains in identification accuracy across age are at odds with the prevailing view that young children use contextual information less effectively than older children, who, in turn, use such information less effectively than adults (Elliott, 1979). Instead, our findings are consistent with the view that 5-year-olds have similar lexical organization to that of older listeners (Cirrin, 1984). Thus, there is every reason to believe that the contextual cues in the test sentences primed similar semantic networks in 5-year-olds, 9-year-olds, and adults.

It is likely that the principal source of discrepancies between the present findings and previous findings (Elliott,

1979; Nittrouer and Boothroyd, 1990) is methodological. Open-set responding (e.g., repeating the final word of test sentences), which is known to pose disproportionate difficulty for young children (Elliott *et al.*, 1979; Geffner *et al.*, 1996), was used in previous studies (e.g., Elliott, 1979), unlike the closed-set responding (picture-pointing) in the present investigation. Moreover, some of the high-predictability sentences in previous research included semantic contextual cues that would be ambiguous or incomprehensible to young children. For example, metaphorical expressions (e.g., ''He played a game of cat and mouse'') such as those used by Elliott (1979) are not mastered until well into the school years (Winner, 1988). Few researchers have provided feedback, which is known to enhance children's performance (Smith and Hodgson, 1970). Most critical, however, is the typical failure to control for age-related differences in stimulus audibility. When all age groups are tested at the same signal-to-noise level, it is difficult to separate the contributions of bottom-up (i.e., sensory) and top-down (i.e., cognitive) factors.

In the present study, reduced latencies to identify words in high-context sentences imply that 5-year-olds, like adults, engage in interpretive processing as each message unfolds. Thus, words appearing early in a sentence constrain listeners' expectations of what words will follow, leading to faster identification of the final words of high-context sentences. Nevertheless, 5-year-olds' response latencies were higher than those of older listeners for low- as well as high-context sentences, a result that can be attributed to faster processing of speech in noise or to more efficient motor processing (i.e., touching the screen) on the part of 9-year-olds and adults.

The addition of 2 dB of noise had comparable effects on performance irrespective of age, disconfirming claims that young children are especially disadvantaged by increases in noise (Mills, 1975). Fallon *et al.* (2000) found that a 5-dB difference in SNR between 5-year-olds and adults—the difference in absolute and masked thresholds between these age groups (Schneider *et al.*, 1986, 1989)—led to comparable word identification accuracy with very simple test sentences (e.g., Touch the dog; Touch the key). Pilot testing with the more complex sentences in the present investigation (e.g., ''Mom talked about the fish''; ''I went to the pond and caught a fish'') indicated that a 6-dB difference in SNR equalized 5-year-olds' and adults' performance in low noise. Ultimately, that SNR difference failed to equalize performance levels across age, especially at the higher noise level. What accounts for this difference between studies? All sentences in Fallon *et al.* (2000) were identical except for the final monosyllabic word, which means that they were equivalent in number of syllables and roughly equivalent in duration. By contrast, sentences in the present investigation varied in their non-target words, number of syllables, overall duration, syntactic structure (''At the soccer game, I waved my flag''), and availability of contextual cues. Such variability may pose differential difficulty for 5-year-olds, just as multiple talkers interfere with preschoolers' ability to identify words in noise (Ryalls and Pisoni, 1997). With increasing age and corresponding growth in cognitive and linguistic ability, children are likely to become more skilled at ignoring salient but irrelevant cues (Bialystock, 1993; Morton and Trehub, 2001).

In contrast to the 12% benefit from context that was evident in the present study, Kalikow *et al.* (1977) reported high-context gains as great as 60% in their adult sample. Note, however, that Kalikow *et al.* (1977) used open-set responding. For adults, Fallon *et al.* (2000) found that the SNR needed to maintain a particular performance level on an open-set task was 13 dB higher than that needed on a closed-set task. Even with low-context sentences, listeners can benefit from cues in the array of pictures. If they hear the sound /b/, for example, they can restrict their choices to the depicted words. Thus, closed-set responding may account for the modest but highly reliable performance gains from semantic cues in the test sentences.

In quiet conditions, 5-year-olds identify the missing target word for high-context sentence stems 99% of the time. Why, then, does performance drop to 75% to 85% when noise is added? The babble background masks the contextual information as well as the target word. Decreasing the perceptibility of semantic cues in the initial portion of the sentence undoubtedly reduces the likelihood that the final word (which is also perceptually degraded) will be identified accurately. What the present investigation indicates is that 5-year-olds are as skilled as older listeners at using perceptually degraded semantic cues within a sentence to enhance their decoding of other perceptual information in that sentence.

Although semantic cues had comparable facilitative effects on 5-year-olds', 9-year-olds', and adults' identification of words in noise, we cannot be sure that comparable processes were operative in each age group. In the real world as opposed to the laboratory, listening is invariably a constructive process. Listeners use their knowledge of the speaker, the topic of discussion, nonverbal cues, and other available information to enhance their decoding of speech. No doubt, constructive listening improves with increasing experience and cognitive sophistication, but the course of development is unknown. Moreover, it remains to be determined whether 5-year-olds would show comparable enhancement when open-set responses are required or when the contextual cues are embedded in more complex discourse.

Obviously, young children will be unable to achieve decoding accuracy comparable to adults in everyday listening situations that do not compensate for sensory differences between children and adults. Our findings indicate, however, that children are able to use strategic listening to their advantage in some situations, at least. One important challenge for future research is to map the developmental course of other aspects of constructive listening.

ACKNOWLEDGMENTS

This research, which was funded by the Canadian Institutes of Health Research, was part of the doctoral dissertation of the first author. We thank Loretta Falco, Michelle Florendo, Salima Ibrahim, and Mark Wellman for assistance in data collection and Kim Yue for technical assistance. We are grateful to Dr. Susan Nittrouer and an anonymous reviewer for their insightful comments on an earlier version of the manuscript.

APPENDIX

	Low context	High context
Practice	He read about the barn.	Farm animals stay in a barn.
	Mom talked about the belt.	I bought Dad a leather belt.
	We looked at the cage.	I put the bird back in its cage.
	He looked at the cheese.	Mice like to eat cheese.
	We pointed at the cloud.	Rain poured from the cloud.
	She read about the cow.	At the farm, I saw a cow.
	Mom pointed at the crown.	The king wore a gold crown.
	He talked about the deer.	In the forest, I saw a deer.
	Dad read about the drum.	Mike banged on a drum.
	She pointed at the fan.	I was hot, so I turned on the fan.
	She talked about the flag.	At the soccer game, I waved my flag.
	We read about the hose.	To water the lawn, Dad used the hose.
	Dad looked at the nose.	The bully punched my nose.
	Dad looked at the phone.	I answered the phone.
	Dad pointed at the pin.	To hold cloth together, we use a pin.
	He pointed at the rose.	I gave my mom a pretty rose.
	We talked about the shell.	At the beach, I found a shell.
	She looked at the snake.	I got bitten by a snake.
	Mom read about the soap.	We wash our hands with soap.
	Mom looked at the sock.	We put the shoe on after the sock.
Experiment	We pointed at the bag.	Mom packed my lunch in a bag.
	Mom pointed at the ball.	We played catch with the ball.
	We read about the bed.	I feel asleep on my bed.
	Dad pointed at the bee.	I got stung by a bee.
	Dad looked at the book.	I like to read a book.
	Mom looked at the boots.	When it snows, I put on my boots.
	We looked at the bread.	Sandwiches are made with bread.
	Mom read about the broom.	I cleaned the floor with a broom.
	Mom pointed at the brush.	To untangle my hair, I use a brush.
	Mom looked at the bus.	Dad rides to work on the bus.
	Mom read about the cake.	For dessert, we ate cake.
	She pointed at the car.	We drove to the store in our car.
	We pointed at the cat.	The dog chased the cat.
	She talked about the chair.	I sat down on the chair.
	She read about the clock.	I knew the time when I looked at the clock.
	He read about the clown.	We laughed at the funny clown.
	Dad pointed at the corn.	Farmers plant rows of corn.
	He talked about the cup.	I drink juice out of a cup.
	We read about the doll.	The girl played with her doll.
	He talked about the door.	Mom asked me to open the door.
	We talked about the dress.	She wore a pretty dress.
	Dad talked about the duck.	At the pond, I fed a duck.
	Mom talked about the fish.	I went to the pond and caught a fish.
	Dad read about the fork.	I eat spaghetti with a fork.
	She pointed at the horse.	I learned how to ride a horse.
	She looked at the house.	Ann's family lives in a house.
	She talked about the key.	To open the door, Dad used a key.
	He looked at the kite.	I like to fly my kite.
	Dad read about the net.	Nick catches bugs with a net.
	Dad looked at the pail.	We carried the water in a pail.
	He read about the pants.	I fell and ripped my pants.
	She pointed at the pig.	The farmer fed the pig.
	He looked at the pot.	Mom cooks dinner in a pot.
	He pointed at the shoe.	I know how to tie a shoe.
	Dad talked about the skunk.	An animal that smells bad is a skunk.
	We looked at the snow.	I like to play in the snow.
	She read about the star.	In the sky, I saw a bright star.
	We talked about the tie.	When Dad gets dressed up, he wears a tie.
	Mom talked about the tree.	A bird built its nest in our tree.
	She looked at the wheel.	My wagon has a broken wheel.

[1]The SNRs yielding 78% performance for low-context sentences are low compared with previous research (e.g., Kalikow *et al.*, 1977; Elliott, 1979). We have demonstrated that the spectral characteristics of the speaker's voice and the closed-set response can account for these unusually low SNRs (Fallon *et al.*, 2000).

Bates, E., Wulfeck, B., Roe, K. Skinner, L., and Zangl, R. (**2001**). ''On-line measures of contextual facilitation in sentence processing by preschool children,'' paper presented at the Society for Research in Child Development, Minneapolis, MN.

Bialystok, E. (**1993**). ''Metalinguistic awareness: The development of children's representations of language,'' in *Systems of Representation in Children: Development and Use*, edited by C. Pratt and A. F. Garton (Wiley, Chichester, UK), pp. 211–233.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). ''Standardization of a test of speech perception in noise,'' J. Speech Hear. Res. **27**, 32–48.

Boothroyd, A., and Nittrouer, S. (**1988**). ''Mathematical treatment of context effects on phoneme and word recognition,'' J. Acoust. Soc. Am. **84**, 101–114.

Cirrin, F. M. (**1984**). ''Lexical search speed in children and adults,'' J. Exp. Child Psychol. **37**, 158–175.

Cole, R. A., and Perfetti, C. A. (**1980**). ''Listening for mispronunciations in a children's story: The use of context by children and adults,'' J. Verbal Learn. Verbal Behav. **19**, 297–315.

Craig, C. H., Kim, B. W., Rhyner, P. M., and Chirillo, T. K. (**1993**). ''Effects of word predictability, child development, and aging on time-gated speech recognition performance,'' J. Speech Hear. Res. **36**, 832–841.

Cycowicz, Y. M., Friedman, D., Rothstein, M., and Snodgrass, J. G. (**1997**). ''Picture naming by young children: Norms for name agreement, familiarity, and visual complexity,'' J. Exp. Child Psychol. **65**, 171–237.

Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. (**2000**). ''Use of context by young and aged adults with normal hearing,'' J. Acoust. Soc. Am. **107**, 538–546.

Dunn, L. M., and Dunn, L. M. (**1997**). *Peabody Picture Vocabulary Test*, 3rd ed. (American Guidance Service, Circle Pines, MN).

Elliott, L. L. (**1979**). ''Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability,'' J. Acoust. Soc. Am. **66**, 651–653.

Elliott, L. L., Connors, S., Kille, E., Levin, S., Ball, K., and Katz, D. (**1979**). ''Children's understanding of monosyllabic nouns in quiet and in noise,'' J. Acoust. Soc. Am. **66**, 12–21.

Fallon, M., Trehub, S. E., and Schneider, B. A. (**2000**). ''Children's perception of speech in multitalker babble,'' J. Acoust. Soc. Am. **108**, 3023–3029.

Fernald, A. (**2001**). ''Making use of semantic context in early language understanding,'' paper presented at the Society for Research in Child Development, Minneapolis, MN.

Fowler, C. A. (**1988**). ''Differential shortening of repeated content words produced in various communicative contexts,'' Lang Speech **31**, 307–319.

Geffner, D., Lucker, J. R., and Koch, W. (**1996**). ''Evaluation of auditory discrimination in children with ADD and without ADD,'' Child Psychiatry Hum. Dev. **26**, 169–180.

Grosjean, F. (**1980**). ''Spoken word recognition processes and the gating paradigm,'' Percept. Psychophys. **28**, 267–283.

Grosjean, F. (**1985**). ''The recognition of words after their acoustic offset: Evidence and implications,'' Percept. Psychophys. **38**, 299–310.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). ''Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability,'' J. Acoust. Soc. Am. **61**, 1337–1351.

Lieberman, P. (**1963**). ''Some effects of semantic and grammatical context on the production and perception of speech,'' Lang Speech **6**, 172–187.

Liu, H., Bates, E., Powell, T., and Wulfeck, B. (**1997**). ''Single-word shadowing and the study of lexical access,'' Appl. Psycholing. **18**, 157–180.

Mills, J. H. (**1975**). ''Noise and children: A review of literature,'' J. Acoust. Soc. Am. **58**, 767–779.

Morton, J. B., and Trehub, S. E. (**2001**). ''Children's understanding of emotion in speech,'' Child Dev. **72**, 834–843.

Nittrouer, S., and Boothroyd, A. (**1990**). ''Context effects in phoneme and word recognition by young children and older adults,'' J. Acoust. Soc. Am. **87**, 2705–2715.

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). ''How young and old adults listen to and remember speech in noise,'' J. Acoust. Soc. Am. **97**, 593–608.

Ryalls, B. O., and Pisoni, D. B. (**1997**). ''The effect of talker variability on word recognition in preschool children,'' Dev. Psychol. **33**, 441–452.

Schneider, B. A., Daneman, M., Murphy, D. R., and Kwong, See, S. (**2000**). ''Listening to discourse in distracting settings: The effects of aging,'' Psychol. Aging **15**, 110–125.

Schneider, B. A., Trehub, S. E., Morrongiello, B. A., and Thorpe, L. A. (**1986**). ''Auditory sensitivity in preschool children,'' J. Acoust. Soc. Am. **79**, 447–452.

Schneider, B. A., Trehub, S. E., and Morrongiello, B. A., and Thorpe, L. A. (**1989**). ''Developmental changes in masked thresholds,'' J. Acoust. Soc. Am. **86**, 1733–1742.

Smith, K. E., and Hodgson, W. R. (**1970**). ''The effects of systematic reinforcement on the speech discrimination responses of normal and hearing-impaired children,'' J. Aud Res. **10**, 110–117.

Snodgrass, J. G., and Vanderwart, M. (**1980**). ''A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity,'' J. Exp. Psychol. Human Learn. Mem. **6**, 174–215.

Studebaker, G. A. (**1985**). ''A 'rationalized' arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.

Tyler, L. K., and Marslen-Wilson, W. D. (**1981**). ''Children's processing of spoken language,'' J. Verbal Learn. Verbal Behav. **20**, 400–416.

Walley, A. C. (**1988**). ''Spoken word recognition by young children and adults,'' Cog. Devel. **3**, 137–165.

Walley, A. C. (**1993**). ''The role of vocabulary development in children's spoken word recognition and segmentation ability,'' Dev. Rev. **13**, 286–350.

Winner, E. (**1988**). *The Point of Words: Children's Understanding of Metaphor and Irony* (Harvard U.P., Cambridge, MA).

The perception of Cantonese lexical tones by early-deafened cochlear implantees

Valter Ciocca,[a] Alexander L. Francis, Rani Aisha, and Lena Wong
Department of Speech & Hearing Sciences, University of Hong Kong, 5/F, Prince Philip Dental Hospital, 34 Hospital Road, Hong Kong SAR, China

(Received 25 July 2001; accepted for publication 26 February 2002)

This study investigated whether cochlear implant users can identify Cantonese lexical tones, which differ primarily in their *F*0 pattern. Seventeen early-deafened deaf children (age = 4 years, 6 months to 8 years, 11 months; postoperative period = 11–41 months) took part in the study. Sixteen children were fitted with the Nucleus 24 cochlear implant system; one child was fitted with a Nucleus 22 implant. Participants completed a 2AFC picture identification task in which they identified one of the six contrastive Cantonese tones produced on the monosyllabic target word /ji/. Each target stimulus represented a concrete object and was presented within a carrier phrase in sentence-medial position. Group performance was significantly above chance for three contrasts. However, the cochlear implant listeners performed much worse than a $6\frac{1}{2}$-year-old, moderately hearing impaired control listener who was tested on the same task. These findings suggest that this group of cochlear implant users had great difficulty in extracting the pitch information needed to accurately identify Cantonese lexical tones. [DOI: 10.1121/1.1471897]

PACS numbers: 43.71.Ky, 43.71.Hw, 43.66.Ts [CWT]

I. INTRODUCTION

Several investigations of pitch perception by cochlear implantees have studied the pitch percepts generated by stimulating the electrodes of multi-channel cochlear implants (see, e.g., Busby *et al.*, 1994; Busby and Clark, 2000; Collins *et al.*, 1997; Nelson *et al.*, 1995; Zwolan *et al.*, 1997). While individual differences among cochlear implantees have been reported in all studies, for a majority of subjects the pitch percepts changed from low to high as the position of the stimulated electrodes moved from the apex to the base of the cochlea in a manner similar to the tonotopic organization of pitch percepts in the normal ear.

For listeners with normal hearing, the pitch of complex sounds (called ''pitch,'' hereafter) is determined mainly on the basis of the frequency of resolved, low-numbered harmonics (Moore *et al.*, 1985; Plomp, 1967; Ritsma, 1967). Dai (2000) demonstrated that harmonics in the vicinity of 600 Hz carry the largest weight in the calculation of pitch for normal-hearing subjects. While the frequencies of resolved harmonics are likely to be the most important cues for pitch perception, it is also possible to obtain a pitch percept, albeit an ambiguous one, from unresolved harmonics (Schouten *et al.*, 1962). Although the presence of a tonotopic organization of pitch percepts is likely to result in the accurate perception of the pitch of pure tones and narrow-band stimuli, it is not well understood how the pitch of complex stimuli is perceived through the electrical stimulation of the cochlea.

The most common processing strategy used by current cochlear implants is the ''continuous interleaved sampling'' (CIS) method (Wilson *et al.*, 1991). This strategy represents complex sounds as a set of amplitude-modulated signals presented through an array of electrodes that are placed within the cochlea. The signals presented through each electrode consist of a carrier pulse that has a 1 to 2 kHz frequency, and whose modulation rate typically preserves temporal information below 400 Hz. Only one electrode is stimulated at any time to prevent interaction of the electrical fields of adjacent electrodes. Faulkner *et al.* (2000) pointed out that listeners who use CIS cochlear implants are not able to resolve low-numbered harmonics of complex sounds whose fundamental frequencies are within the typical range of speech sounds due to the relatively wide bandpass filters used to deliver electrical stimulation to each electrode. Therefore, CIS users should not be able to perceive pitch on the basis of the frequencies of low-numbered, resolved harmonics of complex sounds. CIS implant users might only be able to make use of the weak cues provided by the periodicity information from unresolved harmonics and by overall differences in the amplitude of stimulation across different channels (Geurts and Wouters, 2001). Therefore, cochlear implantees are likely to have difficulties in perceiving the pitch of quasi-periodic sounds like speech and music (Faulkner *et al.*, 2000). Retrieval of within-channel periodicity information should be even more difficult for listeners using cochlear implants which employ low-pulse rate processing strategies such as the SPEAK processing strategy implemented on Nucleus 22 or 24 implants, which employs a pulse rate per channel that is typically lower than 250 Hz (McKay and McDermott, 1993).

While for nontonal languages the availability of auditory cues to pitch perception does not affect the performance on vowel and consonant recognition tasks (Faulkner *et al.*, 2000), different results might be obtained for the perception of languages in which pitch information is used in a contrastive way to cue lexical meaning (Lee and Nusbaum, 1993; Repp and Lin, 1990). For example, in Mandarin each syllable has one of four tones which differ primarily in *F*0 con-

[a]Electronic mail: vciocca@hkusua.hku.hk

tour and level: tone 1 has a relatively high and flat *F*0 contour, tone 2 has a rising contour, tone 3 has a falling and rising contour, and tone 4 has a falling contour. Fu *et al.* (1998) showed that the perception of Mandarin tones can be advantageous for the accurate perception of segmental information. They asked native speakers of Mandarin with normal hearing to perform consonant, vowel, tone, and word recognition tasks by using processed speech. The speech signals were filtered through one, two, three, or four frequency bands. The signal within each band was then half-wave rectified and low-pass filtered at either 50 or 500 Hz in order to remove spectral information within each frequency band while preserving temporal envelope cues. This processing is similar to that of CIS processors, although the most recent versions of the latter employ a larger number of frequency bands (between 8 and 20 bands; see Loizou, 1998, for a review). The results showed that performance improved when the number of frequency bands was increased for the vowel, consonant, and word recognition tasks, but not for the tone recognition task. The advantage of tone recognition in segmental speech perception was shown by the fact that in the one-band, 500-Hz low-pass filtering condition, Mandarin listeners performed better (11% correct) than English listeners (2.9%; Shannon *et al.*, 1995) in a similar word recognition task. In other words, when spectral information is extremely limited, the ability to recognize tones gives an advantage in the recognition of vowels and consonants.

Another important finding of Fu *et al.*'s (1998) study concerns the acoustic cues that can be used for Mandarin tone recognition. They found that performance in tone recognition was well above chance for all conditions, but was affected by the low-pass filtering condition. As expected, the 500-Hz condition produced significantly better tone recognition than the 50-Hz condition. The finding that performance was well above chance even in the 50-Hz condition, for which the temporal envelope cues did not include periodicity information, suggests that listeners might have used temporal envelope cues such as stimulus duration and amplitude contour for identifying tones. This possibility is supported by the finding that both tone and word recognition were highest for tones 3 and 4, for which the *F*0 and the amplitude contours were highly correlated. Listeners were apparently able to recognize these tones with a great degree of accuracy purely on the basis of temporal envelope cues. These results are in agreement with previous findings that Mandarin tones can be recognized on the basis of cues other than *F*0 contour and height, although it is widely recognized that *F*0 contour and height are the main cues to Mandarin tone recognition (Tseng and Massaro, 1986; Whalen and Xu, 1992).

The Cantonese tonal system differs from that of Mandarin in a number of ways. First, Cantonese has six contrastive tones defined according to their pitch height and contour: high level (HL), high rising (HR), mid level (ML), low falling (LF), low rising (LR), and low level (LL). There is also a high falling (HF) tone which does not usually appear in the Cantonese spoken in Hong Kong (Bauer and Benedict, 1997). Second, Cantonese tones have been found to be cued almost exclusively by *F*0 contour and height (Fok Chan, 1974; Vance, 1976). Therefore, Cantonese tones are ideal stimuli for testing the capacity of cochlear implant listeners to estimate the fundamental frequency of phonation for the purpose of perceiving the pitch patterns of speech sounds.

The goal of this study was to investigate the identification of Cantonese tones by early-deafened listeners with cochlear implants. As Busby and Clark (2000) pointed out, the study of early-deafened cochlear implantees can give an insight on the effects of stimulus deprivation on the later development of perceptual skills. In particular, Busby and colleagues found that cochlear implant users who experienced auditory deprivation early in development performed worse on an electrode trajectory discrimination task than implantees who received auditory stimulation at an early age (Busby and Clark, 1996; Busby *et al.*, 1993). They argued that this finding may be related to the increased neural atrophy that results from the lack of stimulation in the developing auditory system. Although this hypothesis would suggest that early-deafened cochlear implantees should have difficulty in the perception of the pitch of complex sounds, the studies by Busby and colleagues used direct electrode stimulation rather than acoustic stimuli to test pitch perception. In the present study, the main goal was to determine whether early-deafened cochlear implantees could extract pitch information from natural speech sounds in order to recognize Cantonese tones.

II. METHOD

A. Subjects

Seventeen native Cantonese-speaking children (nine females, eight males) aged between 4 and 9 years old participated in this study. Only children older than the age of 4 years were included because Ching (1984) showed that even normal-hearing children are unable to reliably recognize isolated lexical tones until age 4. The reported onset of deafness ranged from birth to 30 months; the age at which the implants were fitted varied between 2 years and 6 months and 7 years and 7 months. Early-deafened children using cochlear implants have been found to require at least 12 months experience to have performance level of above 50% accuracy in English lexical stress recognition tasks and close-set word recognition tests (Tyler *et al.*, 1997). Therefore, all listeners had a postsurgical period longer than 12 months except for S1 (11 months), with a range of 11 to 41 months. All children used the Nucleus 24 cochlear implant system (™Cochlear Limited), except for one who used the Nucleus 22 implant. Six employed the SPEAK processing strategy (Seligman and McDermott, 1995), and 11 used Cochlear Limited's ACE speech processing strategy (see Kiefer *et al.*, 2001). For ACE users, the pulse rate was either 1200 Hz (four children), 900 Hz (six children), or 720 Hz (one child). The pulse rate was 250 Hz for all SPEAK users. Participants were fitted with cochlear implants either at the Queen Elizabeth, the Prince of Wales, or the Queen Mary Hospitals in Hong Kong, where they also received auditory and speech training.

B. Stimuli

Natural stimuli were used because children respond best to natural speech tokens in lexical tone identification tasks

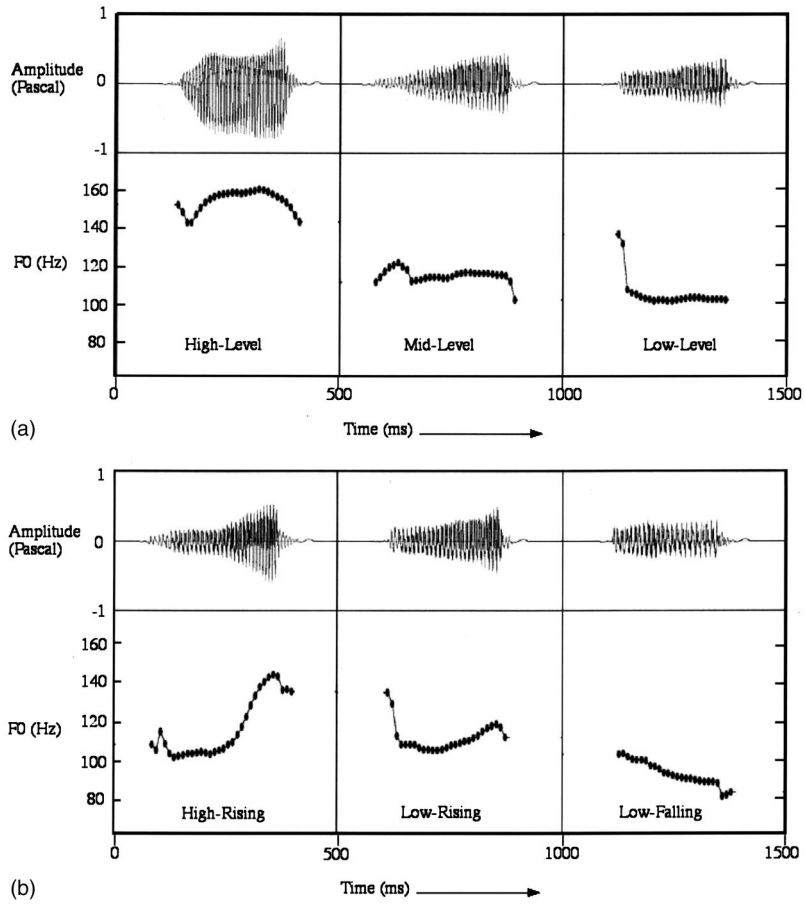


FIG. 1. Amplitude waveforms (top display) and corresponding *F*0 patterns (bottom displays) of the three level tones (a) and of the three contour tones (b) employed in the present study.

(Ching, 1984). The segmental sequence /ji/ was chosen as the basis for the target words as it can be represented by simple and concrete lexical items when produced with any of the six contrastive tones of Cantonese: high-level, /ji55/ (clothing); high-rising, /ji25/ (chair); mid-level, /ji33/ (spaghetti); low-falling, /ji21/ (child); low-rising, /ji23/ (ear); low-level, /ji22/ (two) (see footnote 1). All stimuli were produced by a native Cantonese male speaker aged 21, and recorded at a mouth-to-microphone distance of about 10 cm. The utterances were recorded onto the hard disk of an Apple PowerMacintosh 7100/AV using a Bruel & Kjaer Type 4003 microphone and a Type 2812 MK II microphone preamplifier. The six words were produced ten times each in random order within the carrier phrase:/ŋ⊃23 wui23 t⊃k22 ___ pei25 lei23 teŋ55/ (''I will read ___ for you to hear''). The carrier phrase contained the target words in medial position to mitigate the influence of sentential intonation that might affect the fundamental frequency range of a word in initial or final position (Vance, 1976). The sentence with the smallest total difference from the average *F*0 calculated across all instances of each word was designated as the ''context'' sentence. The productions of each target word with the most extreme *F*0 difference, in the appropriate direction, from the mean *F*0 values of each word were used as the target stimuli. These stimuli were digitally clipped out of their respective sentence, normalized in amplitude, and digitally inserted into the context sentence. Informal listening tests were conducted using native Cantonese listeners with normal hearing to ensure that the carrier sentence with each of the target stimuli sounded natural in terms of prosody and tones.

An acoustic analysis of the fundamental frequency (*F*0) patterns of the target stimuli was conducted with the autocorrelation algorithm of the PRAAT software[2] in order to measure the range of *F*0 variation both within and across stimuli. The results of the *F*0 analyses for each level and rising/falling tones are shown in Figs. 1(a) and (b), respectively, together with the amplitude waveforms for each of the target stimuli. The *F*0 patterns and durations of the target stimuli are typical of Hong Kong Cantonese tones (see Bauer and Benedict, 1997). The mid-level, high-rising, low-rising, low-falling, and low-level had a starting *F*0 of 100 to 110 Hz.[3] Four stimuli had about the same duration; they were the high-level (280 ms), the high-rising (277 ms), the low-rising (283 ms), and the low-falling (269 ms) tones. The mid-level and the high-rising tones had longer duration (336 and 337 ms, respectively). The amplitude envelope is rising over time for the mid-level, low-level, high-rising, and low-rising tones, while it is relatively steady state with short-duration onset and offsets for the high-level and low-falling tones. It is important to stress that for normal-hearing listeners the duration and amplitude envelope cues are not important for the perception of tonal identity in Cantonese (Fok Chan, 1974).

The six target stimuli were grouped into the following eight tonal contrasts: (*i*) high-level versus mid-level (''HL-ML''; tone 55 vs 33), (*ii*) high-level versus low-level (''HL-

LL''; tone 55 vs 22), (*iii*) mid-level versus low-level (''ML-LL''; tone 33 vs 22), (*iv*) high-rising versus low-rising (''HR-LR''; Tone 25 vs 23), (*v*) low-rising versus low-level (''LR-LL''; tone 23 vs 22), (*vi*) low-falling versus low-rising (''LF-LR''; tone 21 vs 23), (*vii*) low-falling versus low-level (''LF-LL''; tone 21 vs 22), (*viii*) high-level versus high-rising (''HL-HR''; tone 55 vs 25). Contrasts HL-ML, HL-LL, and ML-LL were used to investigate the effect of separation between the three pitch levels (high, mid, and low) on tone perception. Pairs HR-LR, LR-LL, LF-LR, and LF-LL were used to test listeners' sensitivity to *F*0 differences in the endpoint of tones, since the tones for these pairs start at similar frequencies but end at different frequencies. Tones in pair HR-LR have the same (rising) contour while tones for pairs LR-LL, LF-LR, and LF-LL have different contours. Finally, pair HL-HR contains tones that have a similar *F*0 endpoint but different initial *F*0.

C. Procedure

Children were tested individually in a double-walled IAC soundproof room. One experimenter sat inside the soundproof room behind the listener, while the another experimenter and the care-giver sat outside the soundproof room. A computer (Power Macintosh 7100/80AV) placed outside the soundproof room and running a Hypercard 2.4 program was used to present the visual and auditory stimuli. Each trial began with the presentation of a target word within the carrier phrase. After this, two pictures of real objects were displayed side-by-side; the pictures represented the two members of a given tonal contrast. Pictures were matched in size (width and height), and in distance from the observer. Visual stimuli were projected from the computer onto a screen placed in the soundproof room using a CTX EzPro 500 projector; projected images were approximately 0.5 m by 0.5 m. The subjects were given the following instructions: ''You will hear each word once, then you should point to one of the two pictures to tell me which word you have heard.'' An experimenter sat outside the soundproof room and recorded the selected response. Participants were encouraged to guess if they were not sure about the correct response. Each contrast was presented four times within a block of trials, twice with each target word. For each target word within a contrast, one trial had the pictures in one order (one picture on the left and the other on the right), the other trial had them in the opposite order. Each participant completed four blocks of trials; each block consisted of 32 tonal contrasts (four trials for each of the eight tonal contrasts). The order of presentation of the stimuli was randomized for each block of trials. The auditory stimuli were output through an Audiomedia II D/A board into a Madsen 0B822 audiometer, and then through a Westra LAB-501 loudspeaker. Participants sat 1 m away from the speaker in the soundproof room.

Before the experimental session, all participants were given 10 to 15 practice trials to ensure that (*i*) they were familiar with all the lexical items and the corresponding pictures, and (*ii*) they understood the nature of the task. These trials were identical to the experimental trials except that the experimenter explained each step of the task, gave feedback, and answered questions. No feedback was provided during the experimental session. Participants were allowed to take a short break whenever they requested it.

The level of the stimuli was measured as peak dBA level at the listening position with a sound-level meter (Bruel & Kjaer, Type 1625). Although each target stimulus was normalized, the peak dBA level for the six target stimuli varied within a range of 8 dBA. The listening level was therefore set such that the range was centered around 65 dBA. At this setting the levels of the target stimuli were the following: 69 dBA for tone 55, 66 dBA for tone 25, 60 dBA for tone 33, 63 dBA for tone 21, 63 dBA for tone 25, and 61 dBA for tone 22. The difference in dBA level among the target stimuli could in principle be a confounding factor in the experiment. However, an informal listening test showed that the perceived loudness of the stimuli was less variable than the dBA readings might suggest. Therefore, it is unlikely that a difference in amplitude level among the stimuli could have been used as a strategy to improve the performance of this task (the results of the experiment fully support this statement).

Before the testing of the cochlear implantees, a moderately hearing impaired child (aged 6 years and 6 months) was used as a control listener. He was wearing a hearing aid and completed a pilot test of the experimental procedure to determine whether the task could be accomplished by a child with hearing impairment (pure tone threshold average) comparable to that of individuals fitted with cochlear implants (Ciocca *et al.*, 2000). The results of this pilot study showed that the moderately hearing impaired listener performed at 92% correct or above on all contrasts except for the HR-LR contrast (58% correct), as expected for a child with normal hearing of the same age (Lui, 2000).

This study was carried out in conjunction with another study (Wong, 2000) that involved tone discrimination and tone identification tasks using other stimuli. Subjects were tested in the order of (a) I1-D-I2 or (b) I2-D-I1, where I1 was the tone identification task of this study while D and I2 were the tone discrimination task and the tone identification task of the other study. Nine subjects were tested in order (a) and eight subjects were tested in order (b).

III. RESULTS

The data were analyzed by computing the percentage of correct scores for each tonal contrast and for each subject. Response rates for individual tones were not calculated because these scores were highly dependent on the choice available to the listeners for a given contrast. For example, the performance for the mid-level tone is likely to be better when it is contrasted with the high-level tone (large *F*0 separation) than when it is paired with the low-level tone (small *F*0 difference).

The average correct scores for each tone contrast ranged from 61% to 50% (see Fig. 2). As a group, the children performed above chance for contrasts HL-ML, HL-LL, and HL-HR (binomial test; $N=272$, $p=\frac{1}{2}$, $\alpha=0.05$). However, even for these contrasts, only a few of the children performed above chance (75% or better) by a binomial test ($N=16$, $p=\frac{1}{2}$, $\alpha=0.05$). Four listeners (S7, S11, S12, and S17) performed above chance for the HL-ML contrast. Only two

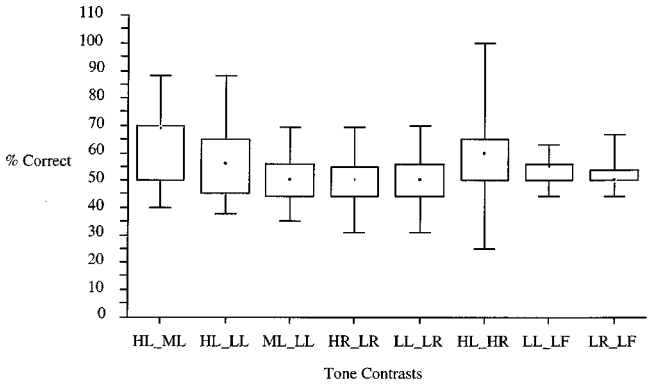


FIG. 2. Box plot of the performance (% correct) of the cochlear implant listeners, showing the median, minimum, and maximum values, and the first and third quartiles for each tonal contrast.

listeners (S12 and S15) performed above chance for contrasts HL-LL and HL-HR. None of the children performed above chance for any of the other five contrasts. Indeed, only 2 out of 17 listeners performed above chance overall (binomial test; $N=128$, $p=\frac{1}{2}$, $\alpha=0.05$).

A one-way ANOVA with repeated measures was carried out on the mean percent correct identification for each subject and each contrast. The results of the ANOVA showed that the means of tonal contrasts were significantly different, $F(7,112)=2.71$, $p<0.05$. Specifically, the performance for contrast HL-ML was significantly better than that for contrasts ML-LL and HR-LR (*post-hoc* Tukey HSD tests, $p<0.05$). None of the other pairwise comparisons between contrasts were statistically significant (Tukey HSD tests, $p>0.05$). These findings suggest that listeners tend to be more accurate at recognizing tones when the alternative choices differed by a large *F*0 separation (HL-ML contrast) than when the *F*0 separation was small (ML-LL) or the *F*0 contours were very similar (HR-LR). An exception to this hypothesis could be the finding that the HL-LL contrast was not perceived with significantly higher accuracy than the ML-LL contrast. However, it is important to notice that performance for the HL-LL, but not the ML-LL, was significantly better than chance.

Given that only two listeners performed above chance overall, it is perhaps not surprising that the correlations between overall performance and age at testing ($r=0.05$, $p>0.05$), duration of the postsurgical period ($r=-0.01$, $p>0.05$), age at implantation ($r=0.1$, $p>0.05$), and onset of deafness ($r=0.1$, $p>0.05$) were not statistically significant. Furthermore, the two listeners who performed best overall (S12 and S15) did not exhibit extreme values that might suggest any trend for any of the above variables.

IV. GENERAL DISCUSSION

Individual results for the tonal contrasts show that very few cochlear implant listeners performed above chance in a tone identification task in which they had to choose between two minimal pair alternatives. As a group, performance was above chance for three out of eight contrasts (HL-ML, HL-LL, and HL-HR), but did not exceed 61% correct on any contrast. These results suggest that early-deafened cochlear implantees have great difficulty in extracting *F*0 information on the basis of the input provided by cochlear implants. The results of a tone discrimination task performed by Wong (2000) using the same listeners further support this suggestion. She presented 30 same/different tone pairs produced by a male speaker in isolation with the syllable/wai/, and found that the overall group performance was 59% correct. Although this performance is above chance by a binomial test ($N=510$, $p=\frac{1}{2}$, $\alpha=0.05$), it does not represent a very accurate performance. Moreover, only four listeners performed significantly above chance in this task by a binomial test ($N=30$, $p=\frac{1}{2}$, $\alpha=0.05$).

Interestingly, all three contrasts that were identified above chance had the high-level tone as one of the members of the pairs. It is possible that group performance in contrast with the high-level tone was better because of the relatively large *F*0 separation between this tone and the other tones. For example, the average *F*0 separation in the level portion of the tones was about 45 Hz between the high-level and the low-level tones, and about 35 Hz between the high-level and the mid-level tones. These separations are well above the *F*0 difference threshold for fundamental frequencies around 150 Hz for CIS implant users (Geurts and Wouters, 2001). On the other hand, the contrast between mid-level and low-level tones was not perceived above chance. These tones were separated over most of their duration by an *F*0 difference (about 10 Hz) which is close to the *F*0 difference threshold for these listeners (Geurts and Wouters, 2001). An alternative reason for the better performance on contrasts involving the high-level tone could be that for some, but not all, speakers this tone has been found to have a higher overall amplitude level than the other tones (Fok Chan, 1974, pp. 139–148). Although this feature of the high-level tone is not produced consistently by all Cantonese speakers, some of the early-deafened children might have learned to exploit this potential cue for the identification of the high-level tone, and they might have used overall amplitude level as a cue for identifying the high-level tone in this experiment. This possibility is supported by the fact that the contrast between low-rising and low-falling tones was not identified above chance. For this contrast, the *F*0 difference at the offset is relatively large (about 40 Hz) but the overall amplitude level of the two tokens is similar. Other potential cues to the identification of tones could be the shape of amplitude envelope and the overall duration. However, Fok Chan (1974) did not identify any amplitude or duration pattern that was consistently associated with specific tones. Although in the current stimuli there were differences in overall amplitude, amplitude envelope, and duration, such differences are not consistently associated with lexical tone differences in the ambient language. Therefore, it is unlikely that the children in this experiment learned to use cues unrelated to pitch to identify Cantonese lexical tones, even though their use might have proven effective with the stimuli employed in this study.

The relatively poor performance of early-deafened cochlear implant listeners could be accounted for by several factors, including the etiology of deafness, the age of the child, the age of implant fitting, and the duration of the postoperative period. However, none of these variables was

found to correlate with identification performance. These results are in apparent contrast with the claim by Busby and Clark (2000) that the duration of auditory deprivation prior to implantation is inversely correlated to performance on an electrode trajectory discrimination task. They found that the implantees in their worst performing group (S11, S15, S17, and S18) also had significantly longer duration of auditory deprivation, and older age at implantation and at testing than the other two groups. However, these group differences do not entirely account for differences in individual performance. For example, their subjects S6, S13, S14, and S16, who belonged to the highest performing group, had a duration of auditory deprivation which was longer (9 years and 5 months or more) than that of their subject S11 (7 years and 6 months). On the other hand, it should also be pointed out that the range of auditory deprivation of the listeners in the present study (1 year and 4 months to 7 years and 1 month) was considerably smaller than that of Busby and Clark's (1 year and 3 months to 18 years and 6 months). Therefore, it is possible that any effects of duration of auditory deprivation would be more difficult to observe in the current study.

The present findings are also in apparent contrast with studies on Mandarin tone perception by cochlear implantees. Huang *et al.* (1995) asked Mandarin-speaking adult implantees who were fitted with the Nucleus 22 implant to perform a four-alternative, forced choice tone recognition task. Their listeners were able to perceive Mandarin tones with about 68% accuracy, compared to a preoperative performance of 34.5% correct. Huang and colleagues suggested that the acoustic cues to fundamental frequency of the four Mandarin tones can be extracted by the speech-coding strategy of the Nucleus 22 cochlear implant system and stimulate the auditory nerve where they are perceived as pitch. There are two differences between the two studies that could account for the seemingly contrasting findings between Huang *et al.*'s and the present study. First, the Mandarin implantees were adults who had auditory and phonological knowledge before the implantation, and therefore might have been able to benefit more from the auditory capacity provided by the implant than early-deafened children (Boothroyd *et al.*, 1991). Adults may also perform better than children in a tone identification task because of more advanced cognitive skills in comparison with the young children. Second, mandarin speakers can make use of temporal envelope cues to recognize lexical tones (Fu *et al.*, 1998; Tseng and Massaro, 1986; Whalen and Xu, 1992). Therefore, the better performance of Mandarin cochlear implantees may have to do with the perception of temporal cues rather then pitch perception. By contrast, it has been shown that such nonpitch cues are unlikely to be reliable cues to tone perception in Cantonese (Fok Chan, 1974; Vance, 1976).

Given that Cantonese cochlear implant listeners are not likely to use temporal envelope cues for tone recognition, their recognition of lexical tones is most probably based on pitch perception. However, the perception of pitch through current multi-channel cochlear implants cannot be accomplished through information about the frequencies of low-numbered, resolved harmonics. Instead, cochlear implantees may rely upon periodicity cues resulting from the interaction of two or more unresolved harmonics. Since periodicity information is generally considered to be a weak cue to pitch, it is perhaps not surprising that early-deafened Cantonese cochlear implant users have difficulties in recognizing lexical tones. Given that even the listeners with higher pulse rates (900 or 1200 Hz) had extreme difficulties in recognizing the lexical tones, it is unlikely than the failure to use periodicity cues for pitch perception was due to a lack of periodicity information transmitted by the processors of the cochlear implants.

Although the performance of this group of cochlear implant listeners was poor overall, it is not possible to know whether these children will be able to improve their Cantonese tone perception skills in the future or whether the relatively impoverished auditory input they receive through the implants will not allow them to learn to identify Cantonese tones in a consistent way. Further studies will have to be carried out in order to determine whether the quality of the auditory input or other cognitive and/or linguistic factors are likely to be the main contributors to the lexical tone perception abilities of Cantonese-speaking cochlear implant listeners.

ACKNOWLEDGMENTS

The authors wish to thank Dr. H.C. Yu and Eddie Wong, ENT Department, Queen Elizabeth Hospital, Hong Kong; Terence Wong, Department of Surgery, Prince of Wales Hospital, Hong Kong; Bessie Pang, Suen Mei Speech and Hearing Center for the Deaf, Hong Kong; Raymond Wu and Donald Chan for technical support; and all participants and their families for their cooperation.

[1]The two digits following each word represent the starting and ending points of the pitch contour of each tone, according to the system presented by Bauer and Benedict (1997), where ''1'' represents the lowest pitch and ''5'' is the highest pitch of a talker's conversational pitch rang. For example, the high-rising tone represented as 25 has a starting point at the lower end of a talker's pitch range (2), and terminates at the highest end of the range (5).

[2]PRAAT 3.9.27 Copyright © 1992–2001 by Paul Boersma and David Weenink (www.praat.org).

[3]The first two pitch pulses of the low-level and low-rising tones had an *F*0 of about 130 Hz. Relatively large variations in *F*0 at the onset and offset of the syllables of tonal languages are not uncommon (see Bauer and Benedict, 1997). However, these short-term variations in *F*0 are typically not perceived as deviations from the overall pitch pattern of the tones. In this case, the low-level and low-rising tones were perceived as having a relatively low starting pitch by native Cantonese listeners with normal hearing.

Bauer, R. S., and Benedict, P. K. (**1997**). *Modern Cantonese Phonology* (Mouton de Gruyter, Berlin).

Boothroyd, A., Geers, A. E., and Moog, J. S. (**1991**). ''Practical implications of cochlear implants in children,'' Ear Hear. **12**, 81–89.

Busby, P. A., and Clark, G. M. (**1996**). ''Electrode discrimination by early-deafened cochlear implant patients,'' Audiology **35**, 8–22.

Busby, P. A., and Clark, G. M. (**2000**). ''Pitch estimation by early-deafened subjects using a multiple-electrode cochlear implant,'' J. Acoust. Soc. Am. **107**, 547–558.

Busby, P. A., Tong, Y.-C., and Clark, G. M. (**1993**). ''Electrode position, repetition rate, and speech perception by early- and late-deafened cochlear implant patients,'' J. Acoust. Soc. Am. **93**, 1058–1067.

Busby, P. A., Whitford, L. A., Blamey, P. J., Richardson, L. M., and Clark, G. M. (**1994**). ''Pitch perception for different modes of stimulation using the Cochlear multiple-electrode prosthesis,'' J. Acoust. Soc. Am. **95**, 2658–2669.

Ching, T. Y. C. (**1984**). ''Lexical tone pattern learning in Cantonese children,'' Lang. Learn. Comm. **3**(3), 317–334.

Ciocca, V., Aisha, R., Francis, A., and Wong, L. (**2000**). ''Can Cantonese children with cochlear implants perceive lexical tones?'' Paper presented at the International Conference on Speech and Language Processing, ICSLP2000, Beijing.

Collins, L. M., Zwolan, T. A., and Wakefield, G. H. (**1997**). ''Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects,'' J. Acoust. Soc. Am. **101**, 440–455.

Dai, H. (**2000**). ''On the relative influence of individual harmonics on pitch judgment,'' J. Acoust. Soc. Am. **107**, 953–959.

Faulkner, A., Rosen, S., and Smith, C. (**2000**). ''Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants,'' J. Acoust. Soc. Am. **108**, 1877–1887.

Fok Chan, Y. Y. (**1974**). *A Perceptual Study of Tones in Cantonese* (Centre of Asian Studies, University of Hong Kong, Hong Kong), Vol. 18.

Fu, Q.-J., Zeng, F.-G., Shannon, R. V., and Soli, S. D. (**1998**). ''Importance of tonal envelope cues in Chinese speech recognition,'' J. Acoust. Soc. Am. **104**, 505–510.

Geurts, L., and Wouters, J. (**2001**). ''Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants,'' J. Acoust. Soc. Am. **109**, 713–726.

Huang, T.-S., Wang, N.-M., and Liu, S.-Y. (**1995**). ''Tone perception of Mandarin-speaking postlingually deaf implantees using the nucleus 22-channel cochlear mini system,'' Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 294–298.

Kiefer, J., Holl, S., Stürzebecher, E., Pfennigdorff, T., and Gstöettner, W. (**2001**). ''Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, ACE) and their relationship to telemetric measures of compound action potential in the nucleus CI 24M cochlear implant system,'' Audiology **40**, 32–42.

Lee, L., and Nusbaum, H. C. (**1993**). ''Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese,'' Percept. Psychophys. **53**, 157–165.

Loizou, P. C. (**1998**). ''Introduction to cochlear implants,'' IEEE Signal Process. Mag. **15, 5**, 101–130.

Lui, J. (**2000**). ''Cantonese tone perception in children,'' unpublished B.Sc. dissertation, University of Hong Kong.

McKay, C. M., and McDermott, H. J. (**1993**). ''Perceptual performance of subjects with cochlear implants using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP),'' Ear Hear. **14**, 350–367.

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**, 1853–60.

Nelson, D. A., Van Tasell, D. J., Schroder, A. C., Soli, S., and Levine, S. (**1995**). ''Electrode ranking of 'place pitch' and speech recognition in electrical hearing,'' J. Acoust. Soc. Am. **98**, 1987–1999.

Plomp, R. (**1967**). ''Pitch of complex tones,'' J. Acoust. Soc. Am. **41**, 1526–1533.

Repp, B. H., and Lin, H.-B. (**1990**). ''Integration of segmental and tonal information in speech perception: A cross-linguistic study,'' J. Phonetics **18**, 481–495.

Ritsma, R. J. (**1967**). ''Frequencies dominant in the perception of the pitch of complex sounds,'' J. Acoust. Soc. Am. **42**, 191–198.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). ''Pitch of the residue,'' J. Acoust. Soc. Am. **34**, 1418–1424.

Seligman, P., and McDermott, H. (**1995**). ''Architecture of the Spectra 22 speech processor,'' Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 139–141.

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

Tseng, C.-Y., and Massaro, D. W. (**1986**). ''Lexical tone perception in Mandarin Chinese: Evaluation and integration of acoustic features,'' in *Linguistics, Psychology, and the Chinese Language*, edited by A. K. a. R. Hoosain (Hong Kong U.P., Hong Kong), pp. 91–104.

Tyler, R., Parkinson, A. J., Bertchy, H. F., Lowder, M. W., Parkinson, W. S., Gantz, B. G., and Kelsay, D. M. (**1997**). ''Speech perception by prelingually deaf children and postlingually deaf adults with cochlear implant,'' Scand. Audiol. Suppl. **26**, (Suppl 46), 65–71.

Vance, T. J. (**1976**). ''An experimental investigation of tone and intonation in Cantonese,'' Phonetica **33**, 368–392.

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

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech recognition with cochlear implant,'' Nature (London) **352**, 236–238.

Wong, A. (**2000**). ''Tone perception performance of prelingually hearing-impaired Cantonese-speaking children with cochlear implants,'' unpublished M.Sc. dissertation, University of Hong Kong.

Zwolan, T. A., Collins, L. M., and Wakefield, G. H. (**1997**). ''Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects,'' J. Acoust. Soc. Am. **102**, 3673–3685.

Representations of sound that are insensitive to spectral filtering and parametrization procedures

David N. Levin[a)]
Department of Radiology, University of Chicago, Chicago, Illinois 60637

(Received 26 June 2001; revised 9 October 2001; accepted 15 February 2002)

This paper describes representations of time-dependent signals that are invariant under any invertible signal distortion. Such a representation can be created by rescaling the signal in a nonlinear dynamic manner that is determined by recently encountered signal levels. Information that is encoded in such representations will be faithfully communicated in the presence of severe signal distortions, which may originate in the transmitter, receiver, or the channel between them. As in speech communication, the receiver is ''blind'' and need not characterize the form of the signal distortion, which remains unknown. The method is applied to analytical examples, acoustic waveforms of human speech, and the short-term Fourier spectra of a bird song. The results suggest that the rescaled representation of a sound is insensitive to the way its spectra have been filtered and parametrized, as long as those processes do not obliterate the differences between the various spectra in the sound. Finally, the possible ''speaker'' independence of these representations is explored in the context of a simple linear prediction model of vocal tracts with a single degree of freedom. [DOI: 10.1121/1.1470164]

PACS numbers: 43.72.Ar, 43.72.−p, 43.72.Ne, 43.60.Lq [DOS]

I. INTRODUCTION

The fidelity of electronic communication is often degraded when the signal is distorted as it propagates through the transmitter, receiver, and the channel between them. Many telecommunications systems attempt to correct for these effects by periodically transmitting calibration data (e.g., test patterns) so that the receiver can characterize the distortion and then compensate for it by ''unwarping'' the signal (e.g., channel equalization). These techniques may be costly because they take the system ''off-line'' for brief periods or otherwise reduce its efficiency.

In contrast, humans perceive the information content of ordinary speech to be remarkably invariant, even though the signal may be transformed by significant alterations of the speaker's voice, the listener's auditory apparatus, and the channel between them.[1–3] Yet, there is no evidence that the speaker and listener exchange calibration data in order to characterize and compensate for these distortions. Evidently, the speech signal is redundant in the sense that listeners extract the same content from multiple acoustic signals that are transformed versions of one another. Human visual perception is also invariant when the raw signal is distorted by a variety of changes in observational conditions. This phenomenon is strikingly illustrated by experiments[4–8] in which subjects wore goggles, creating severe geometric distortions of the visual field (e.g., inversion, reflection, and/or nonlinear warping). Although the subjects initially perceived the distortion, their perceptions of the world returned to the pre-experimental baseline after several weeks of constant exposure to familiar stimuli seen through the goggles. Apparently, humans utilize recent sensory experiences to automatically ''recalibrate'' their perception of subsequent sensory data.

In earlier reports,[9–14] the author showed how to design sensory devices that behave in this way. In such devices, the signal is represented by a nonlinear function of its instantaneous level at each time, with the form of this scale function being determined by the signal levels in a chosen time interval (e.g., recently encountered signal levels). This rescaled signal is invariant if the signal levels at all times are invertibly transformed by the same distortion. This is because the transformation's effect on the signal level at any time is canceled by its effect on the scale function at that time. This can be understood by considering the following analogy. The positions of particles in a plane can be described in terms of a ''natural'' coordinate system (or scale) that is rooted in the particle collection's intrinsic structure; i.e., the coordinate system that originates at the collection's center of ''mass'' and is oriented along its principal moments of ''inertia.'' Each particle's position with respect to this intrinsic scale is invariant under rigid rotations and translations that change all particle coordinates in the extrinsic coordinate system. This is because each particle and the collection's intrinsic coordinate system are rotated and translated in the same manner, so that each particle's location with respect to that coordinate system is unchanged. Earlier papers showed how the signal levels recently detected by a sensory device may have an intrinsic structure that defines a nonlinear coordinate system (or scale) on the manifold of possible signal levels.[9–14] The ''location'' of the currently detected signal level with respect to this intrinsic coordinate system is invariant under any invertible transformation (linear or nonlinear) of the entire signal time series. This is because the signal level at any time and the scale function at the same time point are transformed in a manner that leaves the rescaled signal level unchanged.

In this paper, we show how this self-referential rescaling technique can be used to faithfully communicate information

a)Electronic mail: d-levin@uchicago.edu

in the presence of any invertible signal distortion, without resorting to the explicit calibration procedures that are often used in telecommunications. This is done by encoding the information in the above-described signal invariants, which are not affected by such distortions. The requirement of invertibility is relatively weak; it simply means that the distortion does not compromise the receiver's ability to distinguish between signal levels that are distinguished by the transmitter, and vice versa. In Sec. II, the method is derived, and then it is illustrated with analytic examples. The mathematical properties of the technique are also demonstrated by applying it to the acoustic waveforms of human speech in Sec. III. In Sec. IV, the method is applied to the parametrized spectra of a bird song in order to show that the new sound representations are insensitive to the way the spectra have been filtered and parametrized. In Sec. V, a simple linear prediction model of vocal tracts with a single degree of freedom is used to explore the possible "speaker" independence of the new sound representations. In Sec. VI the implications of this work are discussed, particularly the possible application to speech recognition systems.

II. THEORY

Let $x(t)$ be the time-dependent signal in the transmitter (e.g., the signal driving its antenna circuit), and let X be its value at time T. In this paragraph, we show how to rescale the signal level at this particular time point. The exact same procedure can be used to rescale the signal level at other times, thereby deriving a representation of the entire signal time series. Suppose that $x(t)$ passes through all of the signal levels in $[0,X]$ at one or more times during the interval $T-\Delta T \leqslant t < T$. Here, ΔT is a parameter that can be chosen freely, although it influences the adaptivity and noise sensitivity of the method (see below). At each $y \in [0,X]$, define the value of the function $h(y)$ to be

$$h(y) = \left\langle \frac{dx}{dt} \right\rangle_y , \tag{1}$$

where the right side denotes the derivative averaged over those times in $T-\Delta T \leqslant t < T$ when $x(t)$ passes through the value y. If $h(y)$ is nonvanishing for all $y \in [0,X]$, it can be used to compute the scale function $s(x)$ on this interval

$$s(x) = \int_0^x \frac{dy}{h(y)}. \tag{2}$$

The quantity $S = s(X)$ can be considered to represent the level of the transmitter signal X at time T, after it has been nonlinearly rescaled by means of the function $s(x)$. Now, suppose that the signal in the receiver's detection circuit is related to the signal in the transmitter by the time-independent transformation $x \to x' = x'(x)$. The transformation $x'(x)$ could be the result of a time-independent distortion (linear or nonlinear) that affects the signal as it propagates through the internal circuits of the transmitter and receiver, as well as through the channel between them. Furthermore, suppose that $x \to x'$ is invertible [i.e., $x'(x)$ is monotonic], and suppose that it preserves the null signal [i.e., $x'(0) = 0$]. As mentioned earlier, the requirement of invertibility is relatively weak. It simply means that the distortion does not compromise the receiver's ability to distinguish between signal levels that are distinguished by the transmitter, and vice versa. The transformed signal $x'(t) = x'[x(t)]$ has the value $X' = x'(X)$ at $t = T$. During $T - \Delta T \leqslant t < T$, $x'(t)$ passes through each of the values in $[0,X']$, because of our assumption that $x(t)$ attains all of the values in $[0,X]$ during that time interval. Therefore, for each $y' \in [0,X']$, the process in Eq. (1) can be applied to the transformed signal in order to define the function $h'(y')$ at time T

$$h'(y') = \left\langle \frac{dx'}{dt} \right\rangle_{y'} , \tag{1'}$$

where the right side denotes the derivative averaged over those times in $T - \Delta T \leqslant t < T$ when $x'(t)$ passes through the value y'. By substituting $x'(t) = x'[x(t)]$ in Eq. (1'), using the chain rule of differentiation, and noting that $x(t)$ passes through the value y when $x'(t)$ passes through the value $y' = x'(y)$, we find $h'(y') = dx'/dx|_y h(y)$. The function $h'(y')$ is nonvanishing for $y' \in [0,X']$ because the monotonicity of $x'(x)$ implies $dx'/dx \neq 0$. This means that the process in Eq. (2) can be used to compute a scale function $s'(x')$ on this interval

$$s'(x') = \int_0^{x'} \frac{dy'}{h'(y')}. \tag{2'}$$

The quantity $S' = s'(X')$ represents the level of the receiver signal X' at time T, after it has been rescaled by means of a function $s'(x')$, which was derived from $x'(t)$ just as $s(x)$ was derived from $x(t)$. Because of our assumption that $x = 0$ transforms into $x' = 0$, a change of variables $(y \to y')$ in Eq. (2) implies $s'(x') = s(x)$ and, therefore, $S' = S$. This means that the rescaled value of a signal is invariant under the transformation $x \to x'$. In other words, the rescaled value S of the undistorted signal level at time T, computed from recently encountered undistorted signal levels, will be the same as the rescaled value S' of the distorted signal level at time T, computed from recently encountered distorted signal levels. Now, the above procedure can be followed in order to rescale the signal levels at times other than T. The resulting time series of rescaled signal levels $S(t)$, which the transmitter derives from the transmitted signal $x(t)$ in this way, will be identical to the time series of rescaled signal levels $S'(t)$, which the receiver derives from the received signal $x'(t)$. Thus, if the transmitter encodes information in the rescaled representation $S(t)$ of its signal, that information will be invariantly communicated to the receiver, even in the presence of invertible distortions of the propagating signal.

Notice that the above derivation assumes the prior knowledge that the unknown signal transformation satisfies $x'(0) = 0$. However, a weaker assumption will suffice: namely, prior knowledge of a single pair of "reference" amplitudes (a,a') that are transformed into one another by the unknown transformation: $x'(a) = a'$. In other words, the pair of reference amplitudes does not have to be zero, as was assumed above for simplicity. These reference amplitudes can be used to define dc-offset signal amplitudes: $z = x - a$ and $z' = x' - a'$. Because x and x' are invertibly related and

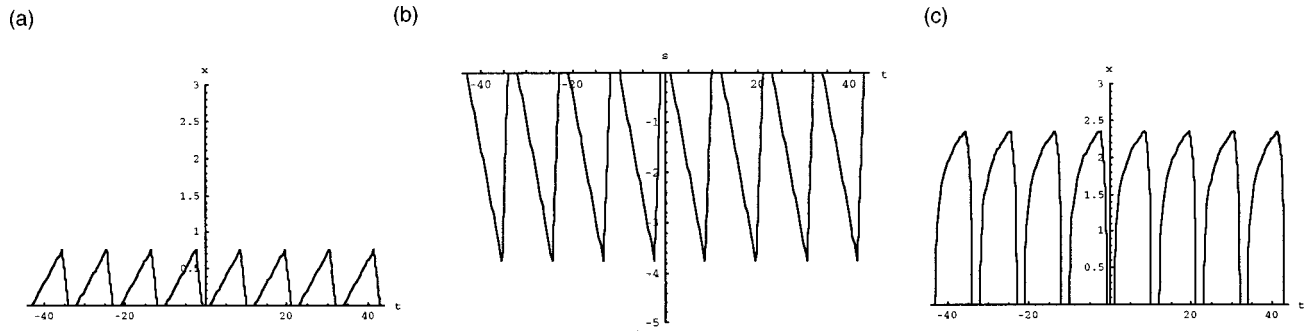


FIG. 1. (a) The transmitter signal $x(t)$ describing the transmission of a long succession of identical pulses that are uniformly spaced in time. (b) The signal representation $S(t)$ that results from applying the rescaling method in Sec. II either to the signal in (a) or to the distorted version of that signal in (c). (c) The receiver signal obtained by subjecting the transmitter signal in (a) to the distortion: $x'(x)=g_1\ln(1+g_2x)$ where $g_1=0.5$ and $g_2=150$.

because the transformations $z(x)$ and $z'(x')$ are invertible, it follows that z and z' are invertibly related. Furthermore, the invertible transformation between z and z' satisfies $z'(0)=0$ because the transformation $x'(x)$ satisfies $x'(a)=a'$. Therefore, the above-described methodology can be directly applied in order to derive invariant representations of $z(t)$ and $z'(t)$.

It is evident that the functional form of a signal's s representation depends on the choice of ΔT. However, the above analytic proof shows that untransformed and transformed signals will have the same s representation for any choice of ΔT, as long as the same value of ΔT is used to derive the s representation of each of these signals. In other words, the choice of ΔT influences the *form* of a signal's s representation, but it does not affect the *invariance* of the s representation for any given choice of ΔT. As demonstrated in Sec. III, the choice of ΔT also influences the manner in which the invariance of the s representation breaks down when the assumptions of the paper are violated (e.g., by time dependence of the unknown signal transformation or by the presence of noise).

Notice that the forms of the scale functions $s(x)$ and $s'(x')$ [and of $h(y)$ and $h'(y')$] will usually be time dependent because they are computed from the time course of previously encountered signals. At some times, both the receiver and transmitter may be unable to compute a rescaled signal level. This will happen if the scale function in Eq. (2) does not exist because the quantity $h(y)$ vanishes for some $y\in[0,X]$ or if the function $h(y)$ cannot even be computed at some values of y because these signal levels were not encountered recently. Because of the monotonicity of $x'(x)$, neither the transmitter nor the receiver can compute a signal invariant at such times. Therefore, if the information is exclusively encoded in signal invariants, no information is transmitted or received at these times. Note that this phenomenon does not result in loss of information or otherwise compromise the fidelity of communication, although it does reduce the time efficiency of the communication process.

It is useful to illustrate these results with a simple example. Suppose the transmitter signal $x(t)$ is a long periodic sequence of triangular shapes, like those in Fig. 1(a). Let a and b be the slopes of the lines on the left and right sides, respectively, of each shape; Fig. 1(a) shows the special case: $a=0.1$ and $b=-0.5$ (measured in inverse time units). If we choose ΔT to be an integral number of periods of $x(t)$, it is easy to see from Eqs. (1) and (2) that the transmitter signal implies $h(y)=(a+b)/2$ and $S(t)=s[x(t)]=2x(t)/(a+b)$ at each point in time. Figure 1(b) shows $S(t)$, which is the transmitted signal after it has been rescaled at each time point as dictated by its earlier time course. Now, suppose that the receiver detects a signal that is distorted by any of the following nonlinear functions: $x'(x)=g_1\ln(1+g_2x)$, where $g_2>0$. For example, if $g_1=0.5$ and $g_2=150$, the distorted signal in the receiver $x'(t)$ looks like Fig. 1(c). When Eq. (1′) is used to compute $h'(y')$ from the received signal, the result is

$$h'(y')=\tfrac{1}{2}(a+b)g_1g_2e^{-y'/g_1} \tag{3}$$

at each point in time. Then, Eq. (2′) shows that the rescaled version of the receiver signal is

$$S'(t)=s'[x'(t)]=\frac{2(e^{x'(t)/g_1}-1)}{g_2(a+b)}. \tag{4}$$

Substituting $x'(t)=x'[x(t)]$ into Eq. (4) shows that $S'(t)=S(t)$. In other words, the rescaled signal $S'(t)$, which the receiver derives from the distorted signal $x'(t)$, is the same as the rescaled signal $S(t)$, which the transmitter derives from the undistorted signal $x(t)$. This is because the effect of the invertible signal transformation on the signal level at any given time $[x(t)\rightarrow x'(t)]$ is *canceled* by its effect on the form of the scale function at that time $[s(x)\rightarrow s'(x')]$. Notice that $s(x)$ and $s'(x')$ [as well as $h(y)$ and $h'(y')$] happen to be time independent in this particular example, and this implies that $x(t)$ and $x'(t)$ are rescaled in a time-independent fashion. This is because, in order to simplify the calculation, $x(t)$ was chosen to be periodic and ΔT was chosen to be an integral number of these periods. In the general case, the scale functions depend on time in a manner dictated by the earlier time course of the signal. However, the transmitter and receiver will still derive identical self-scaled signals [i.e., $S(t)=S'(t)$], as demonstrated by the proof at the beginning of this section and as illustrated by the experimental examples in the next three sections.

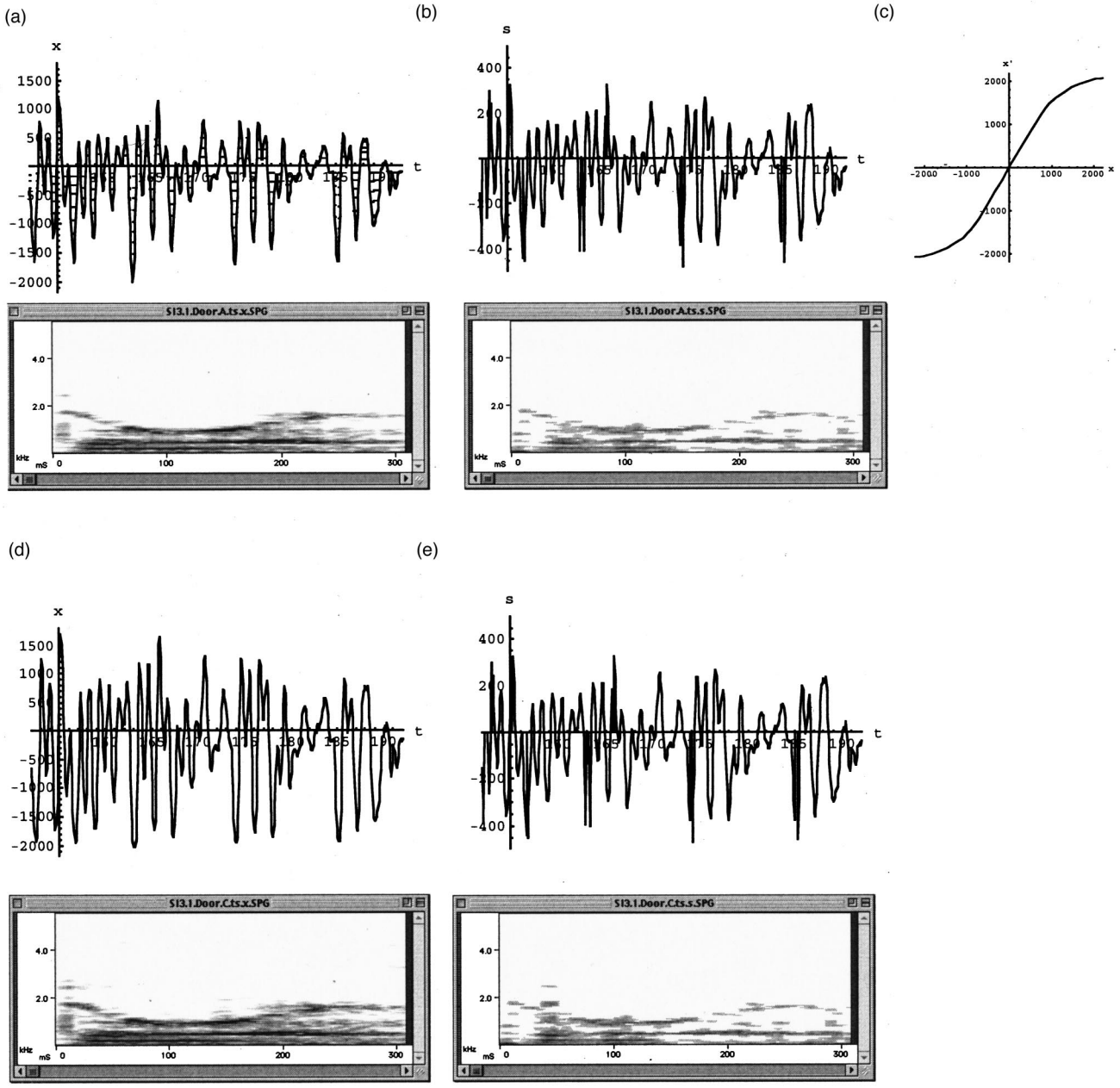


FIG. 2. (a) Upper panel: The digitized acoustic signal of the word ''door,'' uttered by a male speaker of American English. A 40-ms segment in the middle of the 334-ms signal is shown, with time given in ms. The horizontal lines show signal amplitudes that have dynamically rescaled values equal to $s=\pm 50n$ for $n=1,2,\ldots$. Lower panel: The spectrogram of the signal excerpted in the upper panel. (b) Upper panel: the signal $S(t)$ (in units of μs) obtained by rescaling the signal in the upper part of (a) with the parameter $\Delta T=10$ ms. Lower panel: the spectrogram of the rescaled signal excerpted in the upper panel. (c) The nonlinear function $x'(x)$ that was used to transform the signal in (a) into the one in (d). (d) Upper panel: a distorted version of the signal in (a), obtained by applying the nonlinear transformation in (c) to its acoustic waveform. Lower panel: the spectrogram of the signal excerpted in the upper panel. (e) Upper panel: the signal obtained by rescaling the acoustic waveform in (d) with the parameter $\Delta T=10$ ms. Lower panel: the spectrogram of the rescaled signal excerpted in the upper panel.

III. EXPERIMENTS WITH ACOUSTIC WAVEFORMS OF HUMAN SPEECH

In the context of speech recognition systems, the rescaling method should probably be applied to the parametrized short-term Fourier spectra of speech.[11,14] As outlined in Sec. VI, this can be done by generalizing the techniques in Secs. IV and V. However, in this section, we simply apply rescaling to the acoustic waveforms of speech for the purpose of further illustrating the mathematical behavior of the method. The acoustic signals were generated by an adult male American who uttered English words with speed and loudness that were characteristic of normal conversation. These sounds were digitized with 16 bits of precision at a sample rate of 11.025 kHz. Figure 2(a) shows a 40-ms segment of digitized data $[x(t)]$, located at the midpoint of the 334-ms signal corresponding to the word ''door.'' The spectrograms in this figure and in the rest of this section were produced by computing the short-term Fourier transform of the acoustic waveform in 256-sample Hamming windows that were centered at 64-sample intervals. Figure 2(b) shows the ''s representa-

tion'' [i.e., the rescaled signal $S(t)$] that was derived from Fig. 2(a) by the method of Sec. II. The value of S was determined at each time point by a scale function $s(x)$, which was derived from the previous 10 ms of signal (i.e., $\Delta T = 10$ ms). These scale functions are shown by the horizontal lines in Fig. 2(a), which denote values of x corresponding to $s = \pm 50n$ for $n = 1,2,\ldots$. Figure 2(d) shows the signal that was derived from Fig. 2(a) by means of the nonlinear transformation $[x'(x)]$ shown in Fig. 2(c). For example, this could represent the effect of a nonlinear microphone or amplifier. Figure 2(e) is the rescaled signal that was derived from Fig. 2(d) with the parameter ΔT chosen to be 10 ms. Although there are significant differences between the ''raw'' signals in Figs. 2(a) and (d), their s representations [Figs. 2(b) and (e)] are almost identical, except for a few small discrepancies that can be attributed to the discrete methods used to compute derivatives. Thus, the s representation was invariant under a nonlinear signal distortion, as expected from the derivation in Sec. II. It is interesting to note that this result is apparent when one listens to the sounds represented in Fig. 2 (see the sounds posted at the web site in Ref. 15). Although all four signals in Fig. 2 sound like the word ''door,'' there is a clear difference between the sounds of the two raw signals, and there is no perceptible difference between the sounds of their rescaled representations. In general, the rescaled signals sound like the word door, uttered by a voice degraded by slight ''static.''

Some comments should be made about technical aspects of the example in Fig. 2. The rescaled signals in Figs. 2(b) and (e) were computed by a minor variant of the method in Sec. II. Specifically, we assumed that all signal distortions were *monotonically positive*, and we restricted the contributions to Eq. (1) and Eq. (1′) to those time points at which the signal had a *positive* time derivative as it passed through the values y and y', respectively. The rescaled signal is still invariant because monotonically positive transformations do not change the sign of the signal's time derivative, and, therefore, the functions $h(y)$ and $h'(y')$ were still constructed from time derivatives at identical collections of time points. At each time point, we attempted to compute the rescaled signal from the signal time derivatives encountered during the most recent 10 ms ($\Delta T = 10$ ms). At some times, the signal could not be rescaled because the signal level at that time was not attained during the previous 10 ms, and, therefore, there were no contributions to the right side of Eq. (1) for some values of y. For example, this happened at $t \sim 163$, 174, and 185 ms in Fig. 2. At such times, a signal invariant could not be computed, and communication of distortion-invariant information was not possible. As mentioned in Sec. II, this occurs at identical time points when rescaling is applied to the ''undistorted'' signal [e.g., Fig. 2(a)] and to any distorted version of it [e.g., Fig. 2(d)]. This means that the s representations of all of these signals are nonexistent at identical time points and that at all other times they exist and have the same values. Therefore, this phenomenon does not corrupt the invariance of the signal's s representation, although it does reduce its information content. In this experiment, the s representation could not be computed at 8% of all time points. The rescaled signal was set equal to zero at those times.

Figure 3 shows what happened when the nature of the distortion changed abruptly. The signal in Fig. 3(b) was derived by applying the nonlinear transformation in Fig. 2(c) to the first half (i.e., the first 167 ms) of the signal excerpted in Fig. 2(a) and by applying the nonlinear transformation in Fig. 3(a) to the second half of that signal. Figure 3(c) shows the s representation derived by dynamically rescaling Fig. 3(b) with $\Delta T = 10$. Comparison of the latter to Fig. 2(b) shows that the s representation was invariant except during the time period 167 ms$\leqslant t \leqslant$177 ms. These discrepancies can be understood in the following way. During this time interval, the rescaled signal in Fig. 3(c) was derived from a mixed collection of signal levels, some of which were transformed as in Fig. 2(c) and some of which were transformed as in Fig. 3(a). This violates the proof of invariance (Sec. II), which assumed the time independence of the transformation between the ''undistorted'' and ''distorted'' signals. Notice the transitory nature of this corruption of the s representation. The rescaled signals in Figs. 2(b) and 3(c) became identical again, once sufficient time (ΔT) elapsed for the distortion to become constant over the time interval utilized by the rescaling procedure. In other words, the dynamic rescaling process was able to adapt to the new form of the distortion and thereby ''recover'' from the disturbance. Therefore, if communicating systems are encoding information in the signal's s representation, faithful communication will be reestablished ΔT time units after the onset of a change in the transmitter or the receiver or the channel between them. This adaptive behavior resembles that of the human subjects of the goggle experiments mentioned in Sec. I.

Figure 4 illustrates the effect of noise on dynamic rescaling. Figure 4(a) was derived from Fig. 2(d) by adding white noise so that the signal-to-noise ratio was equal to 15 dB. This causes a pronounced hiss to be superposed on the word door when the entire 334-ms sound exemplified by Fig. 4(a) is played (see the sounds posted at the web site in Ref. 15). Figure 4(b) is the s representation, derived by dynamically rescaling Fig. 4(a) with $\Delta T = 10$ ms. Comparison of Figs. 4(b), 2(e), and 2(b) shows that the noise has caused some degradation of the invariance of the s representation. This is expected because additive noise ruins the invertibility of the transformations relating Figs. 4(a), 2(d), and 2(a), thereby violating the proof of the invariance of S in Sec. II. The noise sensitivity of the s representation can be decreased by increasing ΔT, because this increases the number of contributions to the right side of Eq. (1), which tends to ''average out'' the effects of noise. However, such an increase in ΔT means that more time is required for the dynamic rescaling process to adapt to a sudden change in distortion.

IV. EXPERIMENTS WITH THE SPECTRA OF A BIRD SONG

In this section, we consider a bird song whose short-term Fourier spectra oscillated along a curve in the multidimensional space of spectral parameters. At any given time, the sound's spectrum was characterized by $x(t)$, a function that gave its position along this curve. We demonstrate that

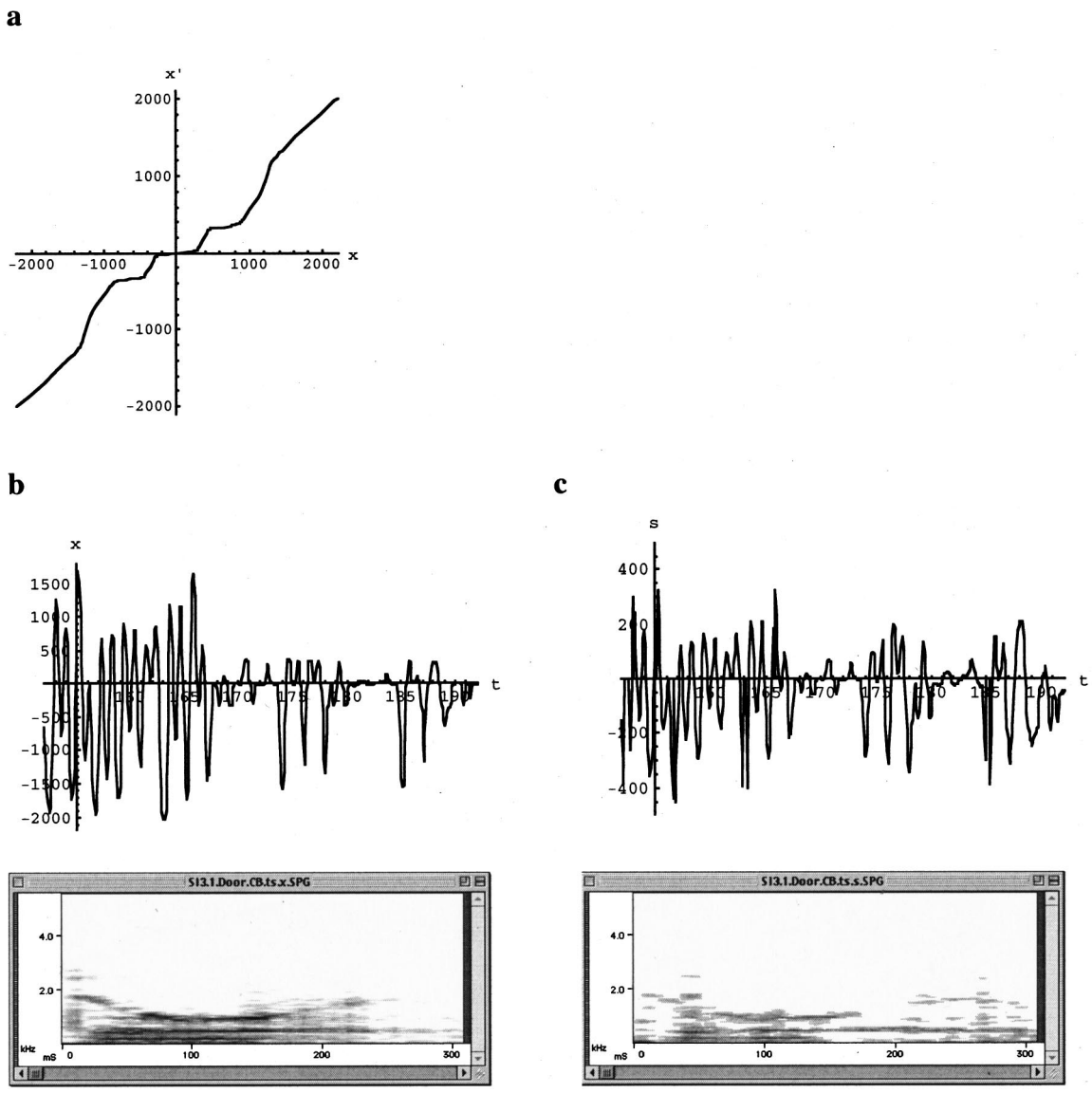


FIG. 3. This figure shows the effects of an abrupt change in the distortion of the signal. (a) A nonlinear signal distortion. (b) Upper panel: the signal obtained by applying the transformation in Fig. 2(c) to the first half (167 ms) of the acoustic waveform excerpted in Fig. 2(a) and by applying the distortion in panel (a) to the second half of that waveform. Lower panel: the spectrogram of the distorted acoustic waveform excerpted in the upper panel. (c) Upper panel: the signal obtained by rescaling the acoustic waveform in panel (b) using the parameter $\Delta T = 10$ ms. Lower panel: the spectrogram of the rescaled signal excerpted in the upper panel.

this function has a self-referential representation $S(t)$, which is unaffected by passing the sound through a wide variety of filters. We also demonstrate the insensitivity of this self-referential representation to the method of parametrizing the sound's spectra.

We considered a 260-ms sound of a warbler that consisted of one long ''chirp'' and two short ''chirps'' (CANARY: The Cornell Bioacoustics Workstation, Version 1.2, Cornell Laboratory of Ornithology, Ithaca, NY). This sound was digitized at 22.3 kHz with 16 bits of precision and subjected to a short-term Fourier transform, after it had been ''windowed'' with a Hamming function in 128-sample frames, which were centered at 32 sample intervals. The time-domain waveform and spectrogram of the sound are shown in Fig. 5(a). The magnitude of each spectrum was subjected to a discrete cosine transformation (DCT), after it had been smoothed by averaging over bins of 800 Hz width. Thus, each spectrum was represented by a single point in the 64-dimensional space of DCT coefficients. Figure 5(b) displays these points after they were projected onto the three-dimensional subspace of the coefficients with indices $i = 1$, 4, and 7. It is evident from this figure (as well as from the points' positions in other 3D subspaces) that all of the sound's spectra lie close to an arc-like curve in the space of all DCT coefficients. During the first chirp, the spectrum moved from a position at the densely populated end of the curve to its sparsely populated distal segment and then followed the reverse trajectory back to its initial position. During each of the second and third chirps, the spectrum made oscillatory movements along the curve's proximal segment, before coming to rest near its initial position. We considered a simulated sound detector that ''sensed'' the nine spectral DCT coefficients with indices $i = 1,4,7,10,12,13,15,16,18$. The detector's sensor state $x(t)$ consisted of the spectrum's

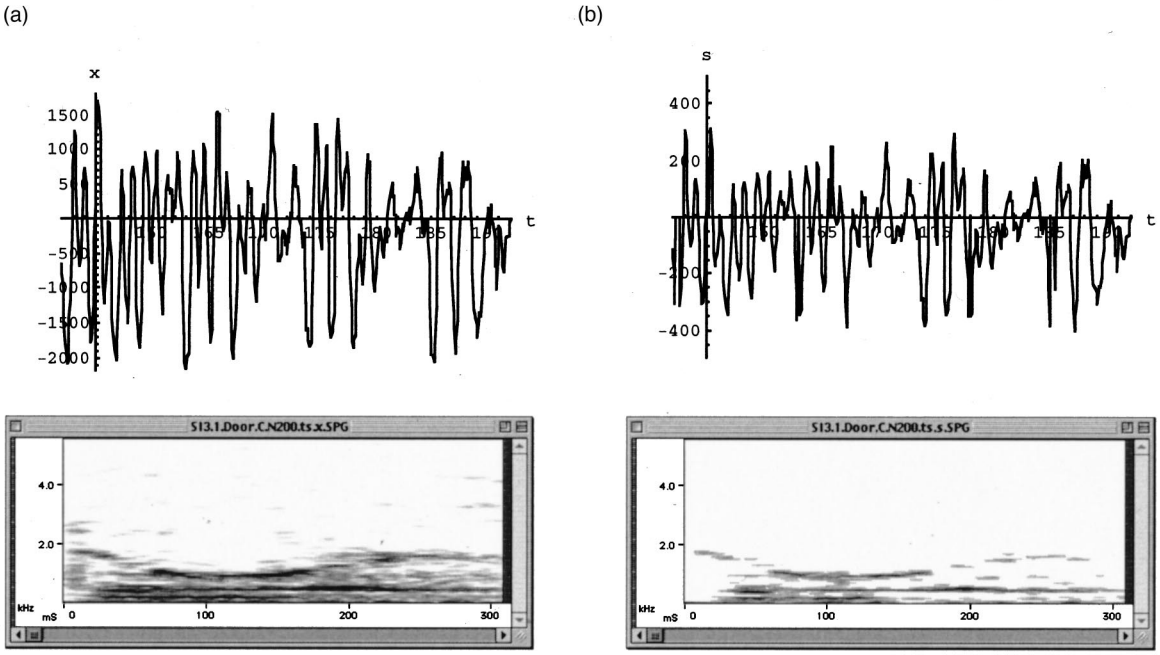


FIG. 4. The effect of noise on the rescaling process. (a) Upper panel: the signal derived from the signal in Fig. 2(d) by adding white noise so that the signal-to-noise ratio is 15 dB. Lower panel: the spectrogram of the entire time course of the signal excerpted in the upper panel. (b) Upper panel: the signal obtained by rescaling the signal in panel (a) with $\Delta T = 10$ ms. Lower panel: the spectrogram of the entire time course of the signal excerpted in the upper panel.

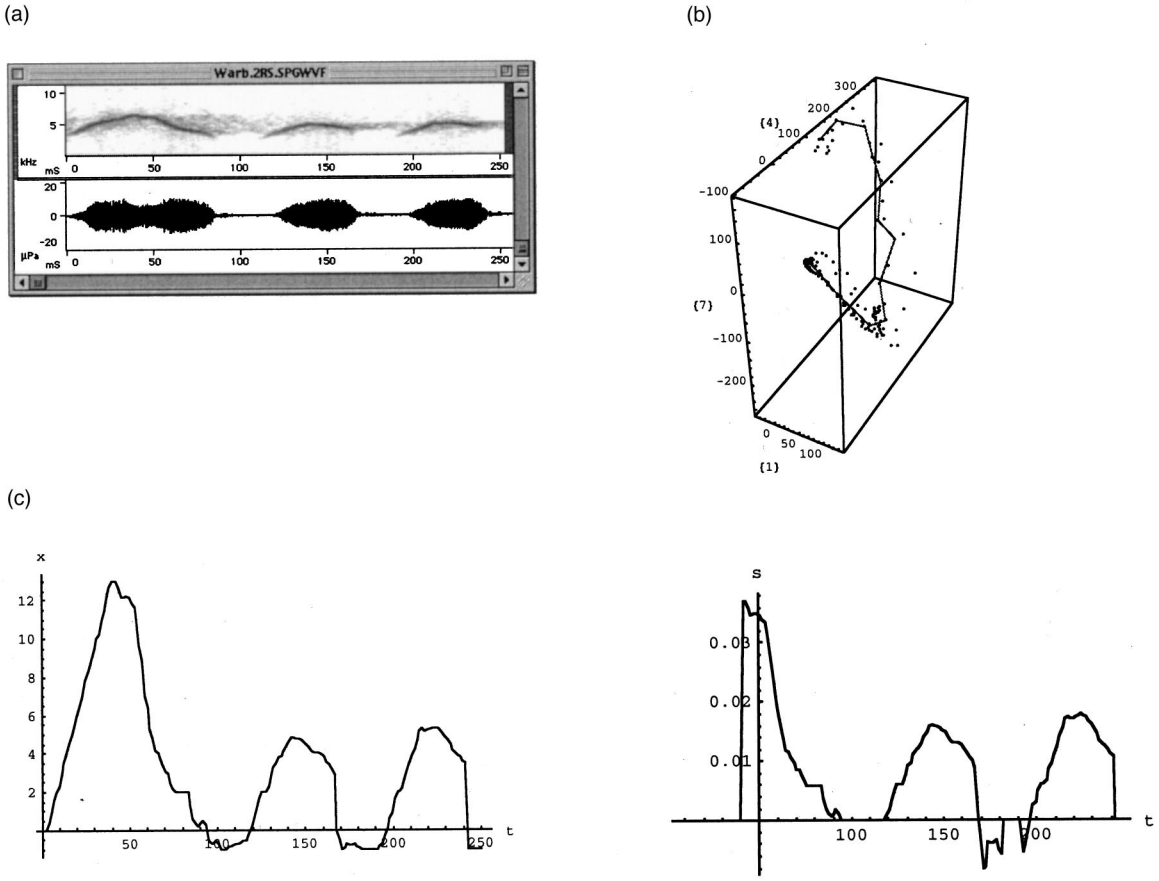


FIG. 5. (a) The spectrogram (upper) and acoustic waveform (lower) of a warbler's song. (b) The positions of the spectra in (a) in the space of the three spectral DCT coefficients with indices $i = 1,4,7$. The solid lines depict a piecewise linear curve that is close to the nearly one-dimensional collection of points. (c) Left panel: the sensor state of a detector that measured the position of each sound spectrum along a curve in the space defined by the nine spectral DCT coefficients with indices $i = 1,4,7,10,12,13,15,16,18$. Right panel: the rescaled representation of the sensor data on the left.

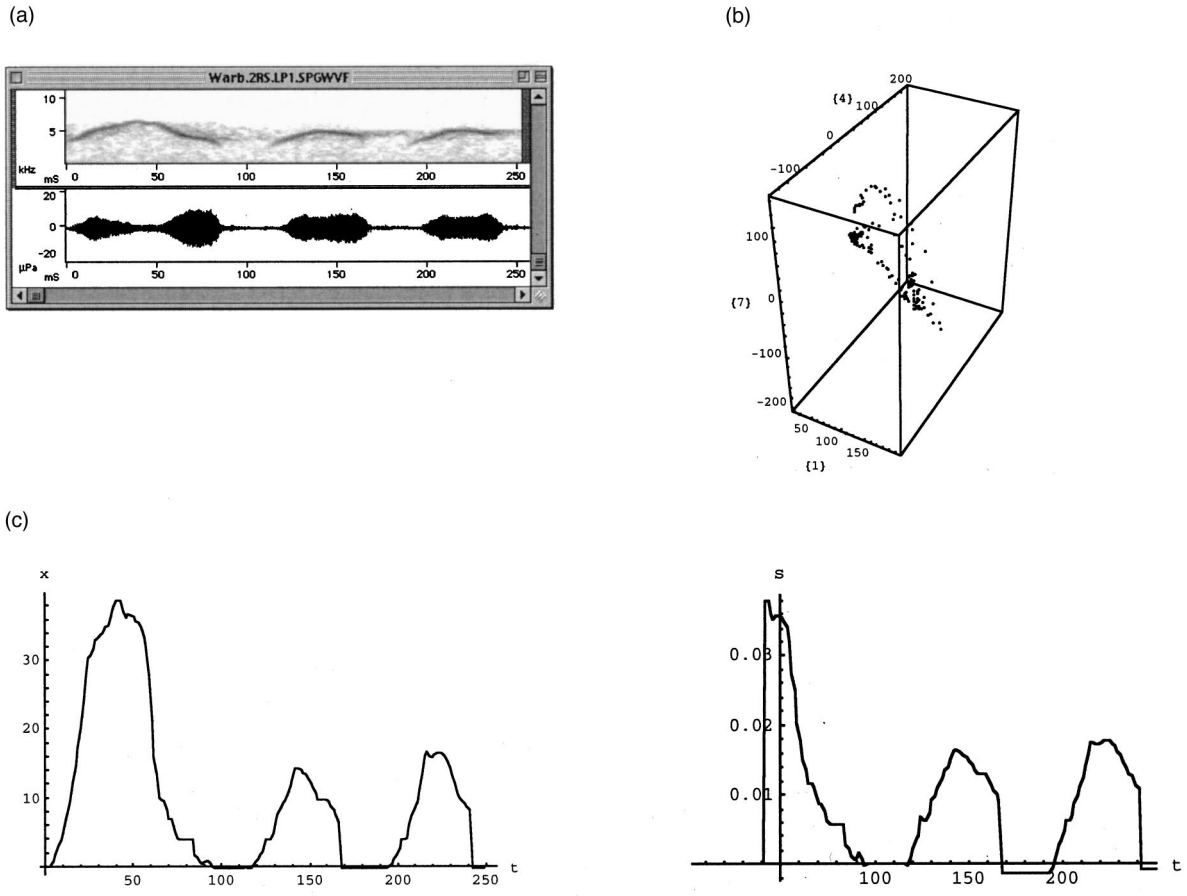


FIG. 6. The effect of filtering out high-frequency components of the sound in Fig. 5(a). (a) The spectrogram (upper) and acoustic waveform (lower) of the warbler's song in Fig. 5(a), after it was passed through a Hamming filter (1.5 ± 5.6 kHz). (b) The positions of the spectra in (a) in the space of the spectral DCT coefficients with indices $i = 1,4,7$. (c) Left panel: the sensor state of a detector that measured the position of each sound spectrum along a curve in the space defined by the nine spectral DCT coefficients with indices $i = 1,4,7,10,12,13,15,16,18$. Right: the rescaled representation of the sensor data on the left.

position along the above-described curve in this nine-dimensional subspace. There are many dimensional reduction procedures that can be used to assign each spectral point to a position on this one-dimensional manifold. In this example, this was done by defining a piecewise linear trajectory that hugged the curvilinear collection of spectral points in the nine-dimensional subspace [Fig. 5(b)], and each spectrum was assigned the position (x) of the nearest point on this collection of chords. The x coordinate system along these chords was defined so that its origin was at the position of the first spectrum in the sound [i.e., it was defined so that $x(0) = 0$]. The left panel of Fig. 5(c) shows the spectrum's position as a function of time, which constitutes the detector sensor state elicited by the sound in Fig. 5(a). The right panel of Fig. 5(c) shows the function $S(t)$ obtained by using Eqs. (1) and (2) to rescale the left panel, with the choice $\Delta T = 300$ ms. Notice that the rescaling process has segmented the sound into the three chirps, separated by time intervals during which invariants could not be computed. As proved in Sec. II, the same rescaled function $S(t)$ would have been computed if we had used *any* x coordinate system along the sound's trajectory, as long as it was invertibly related to the above-described scale and originated at the same point on the curve. For instance, we would have obtained the same rescaled sensor state if we had projected the spectral points onto *any* piecewise continuous collection of chords that was sufficiently close to the arc-like collection of spectral points.

We attenuated the high-frequency components of the sound in Fig. 5 by convolving it with a Hamming function corresponding to a filter passing frequencies in the range 1.5 ± 5.6 kHz. Comparison of Figs. 6(a) and 5(a) shows the suppression of the high frequencies in this sound, which had a noticeably lower ''pitch'' than the one in Fig. 5(a) (see the sounds posted at the web site in Ref. 15). As before, the DCT coefficients of the sound's spectra defined points that tended to cluster along an arc-like line in 64-dimensional DCT space. Figure 6(b) shows these points after they were projected onto the subspace defined by the DCT coefficients with indices $i = 1$, 4, and 7. The left panel of Fig. 6(c) shows each spectrum's position along the corresponding curve through the nine-dimensional subspace of DCT coefficients ''sensed'' by the simulated detector (i.e., coefficients with indices $i = 1,4,7,10,12,13,15,16,18$). This function represents the detector's sensor state $x(t)$ as it ''listened'' to the filtered sound in Fig. 6(a). The right panel of Fig. 6(c) shows the function $S(t)$ that was obtained by rescaling the left panel with the choice $\Delta T = 300$ ms. Figures 5(c) and 6(c) demonstrate that the original and filtered sounds had similar rescaled representations, even though they produced significantly different sensor states in the simulated detector. This

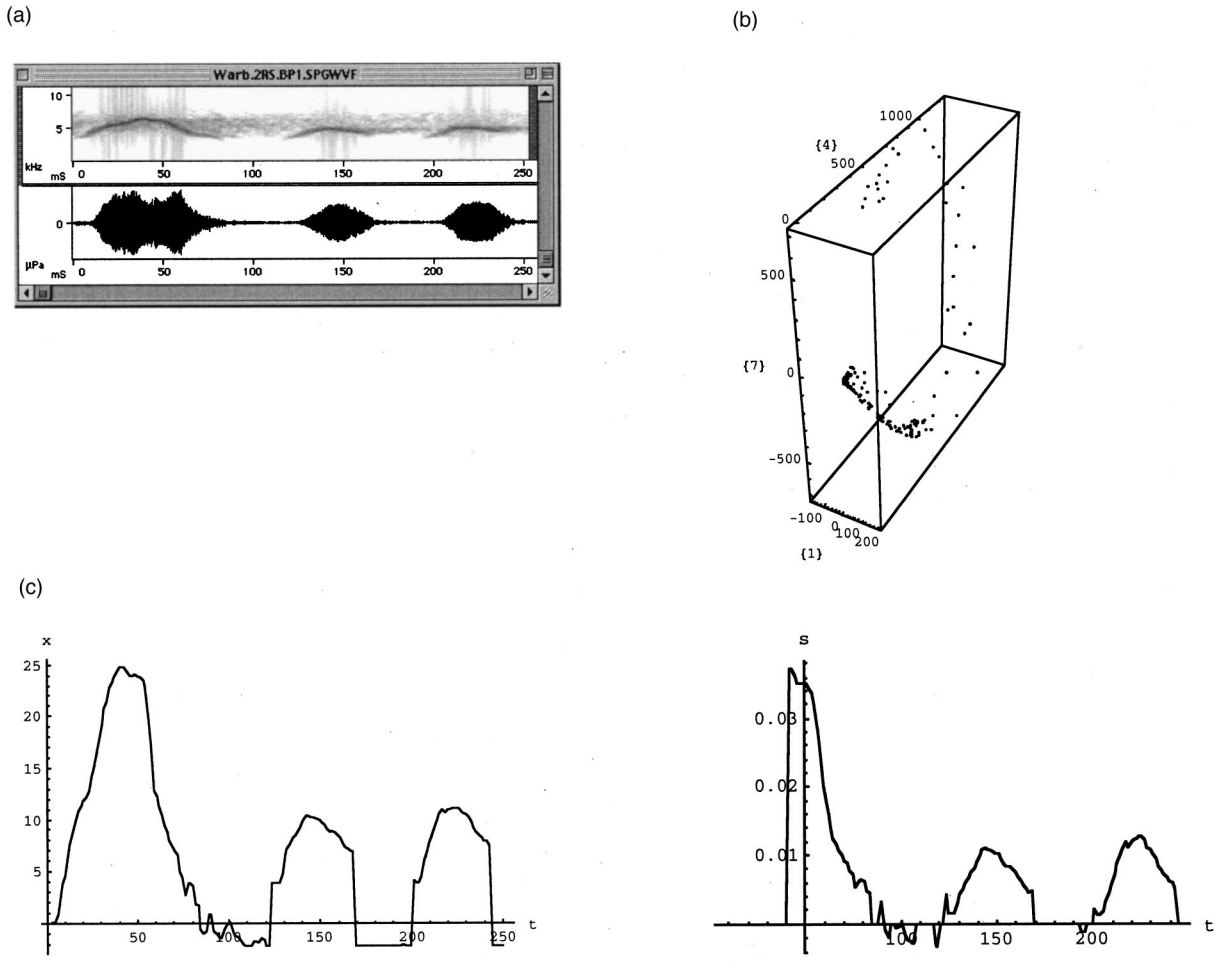


FIG. 7. The effect of filtering out low-frequency components of the sound in Fig. 5(a). (a) The spectrogram (upper) and acoustic waveform (lower) of the warbler's song in Fig. 5(a), after it was passed through a Hamming filter (6.0 ± 2.8 kHz). (b) The positions of the spectra in (a) in the space of the spectral DCT coefficients with indices $i = 1,4,7$. (c) Left panel: the sensor state of a detector that measured the position of each sound spectrum along a curve in the space defined by the nine spectral DCT coefficients with indices $i = 1,4,7,10,12,13,15,16,18$. Right: the rescaled representation of the sensor data on the left.

is due to the fact that the unrescaled sensor states were related by an invertible transformation that preserved null values. Such a transformation existed because two conditions were satisfied: (1) an invertible transformation related each unfiltered spectrum's position on the one-dimensional manifold of the sound's spectra to the corresponding filtered spectrum's position on the one-dimensional manifold of spectra in the filtered sound; (2) for both the unfiltered and filtered sounds, the position of each spectrum on the one-dimensional manifold of the sound's spectra was invertibly related to its coordinate along the sound's curve in the subspace of ''sensed'' DCT coefficients. The first condition was satisfied because the filtering operation did not obliterate all of the differences between any pair of spectra in the sound; i.e., it did not map different spectra in the original sound onto the same filtered spectrum. Specifically, because the width of the filter's impulse response was less than the width of the window used to create short-term spectra, at every time point each component of the filtered spectrum was equal to the corresponding component of the unfiltered spectrum, multiplied by the corresponding spectral component of the filter's impulse response. Because the filter's spectrum was nonzero for most important frequencies, this transformation did not map two different spectra in the unfiltered sound onto a single spectrum of the filtered sound. The second condition means that the ''sensed'' DCT coefficients were sensitive to the differences among the sound's spectra. This was true because the spectra followed a trajectory that was never simultaneously orthogonal to the axes of all of the sensed coefficients. Because the above two conditions were satisfied in this example, there was an invertible transformation between the sensor states derived from unfiltered and filtered spectra at identical times. Furthermore, this mapping preserved the null value. This is because the origins of the x coordinate systems for the two sounds (unfiltered and filtered) were both defined to be at the positions of the first spectrum in each sound, and the mapping transforms these positions into one another. In other words, in this example, two detectors, which sensed a sound through unfiltered and filtered channels, used the first detected spectrum to define a common origin for their sensor state scales. This procedure is analogous to having a choir leader play a pitch pipe in order to establish a common origin of the musical scale among the singers.

Figure 7(a) shows the sound produced by attenuating the low-frequency components of the sound in Fig. 5(a). This

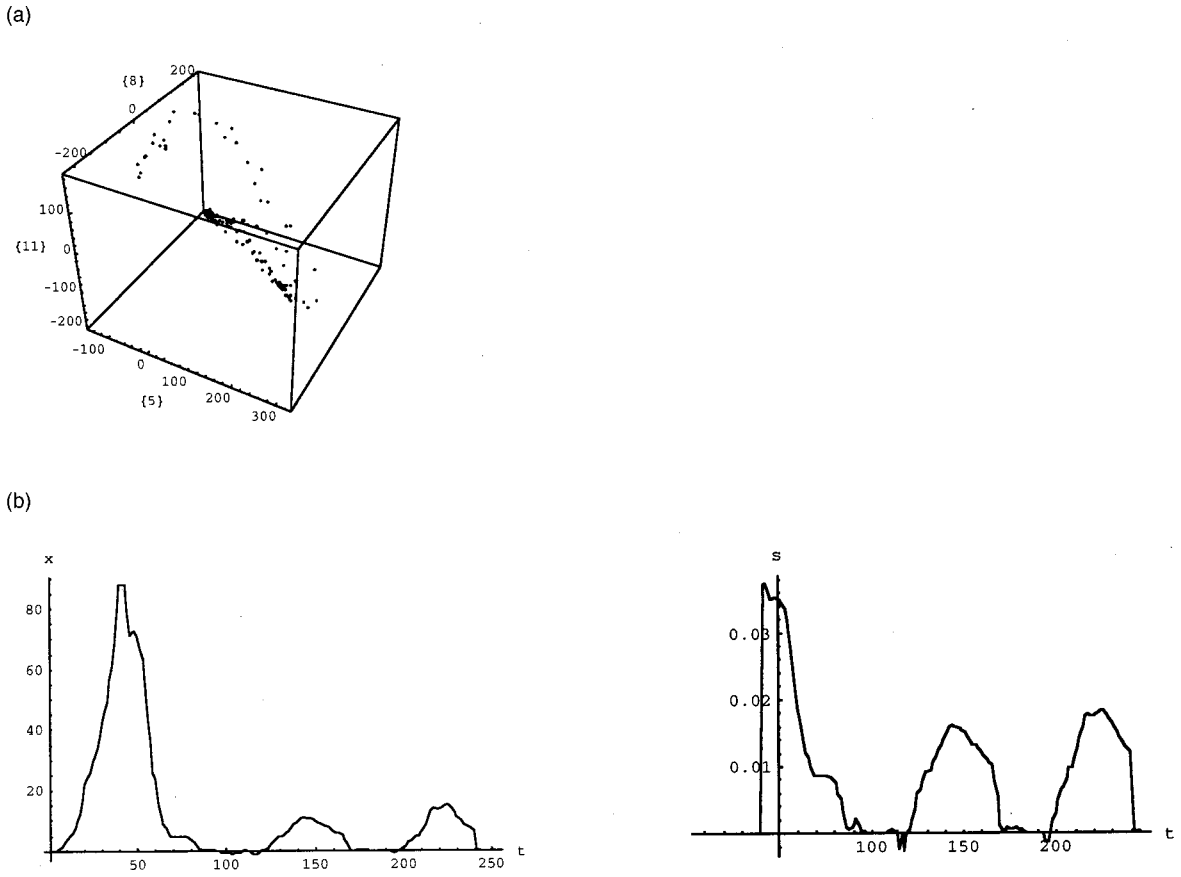


FIG. 8. The effect of using a different simulated detector to sense the sound in Fig. 5(a). (a) The positions of the spectra in Fig. 5(a) in the space of the spectral DCT coefficients with indices $i=5,8,11$. (b) Left panel: the sensor state of a detector that measured the position of each sound spectrum along a curve in the space defined by the eight spectral DCT coefficients with indices $i=5,6,8,9,11,14,17,19$. Right: the rescaled representation of the sensor data on the left.

sound was created by convolving the sound in Fig. 5(a) with a Hamming function corresponding to a filter passing frequencies in the range 6.0 ± 2.8 kHz. Comparison of Figs. 7(a) and 5(a) shows the suppression of the low frequencies of this sound, which had a high-pitched quality, noticeably different from the sound in Fig. 5(a) (see the sounds posted at the web site in Ref. 15). Figure 7(b) shows three DCT coefficients (indices $i=1$, 4, and 7) of the sound's spectra, which followed a curvilinear trajectory in the full DCT coefficient space, as well as in the depicted three-dimensional subspace. The left panel of Fig. 7(c) shows the detector's sensor state $x(t)$ as it listened to the filtered sound in Fig. 7(a); namely, the figure shows each spectrum's position along the sound's curve through the subspace of the nine sensed DCT coefficients (indices $i=1,4,7,10,12,13,15,16,18$). The right panel of Fig. 7(c) shows the function $S(t)$ that was obtained by rescaling the left panel with the choice $\Delta T=300$ ms. It is evident from Figs. 5(c) and 7(c) that similar rescaled representations were derived from the original and filtered sounds, even though they produced noticeably different unrescaled sensor states. As before, this is expected because the sensor-state time series produced by the two sounds were related by an invertible transformation. Notice that this transformation was nonlinear even though the spectral coefficients of the two sounds were related by a linear filtering operation. This is because the transformation between sensor states was a composite of mappings that related the positions of spectra in nonlinear subspaces (i.e., arc-like curves) in the spaces of spectral coefficients.

Next, consider what happened when the sound in Fig. 5(a) was sensed by a simulated detector with different ''sensors.'' This detector sensed eight DCT coefficients (with indices $i=5,6,8,9,11,14,17,19$) that were completely different than those recorded by the previously described simulated detector. The sound's spectra were represented by points in the eight-dimensional space of these coefficients, and these points were on (or nearly on) a curve. This is illustrated by Fig. 8(a), which depicts the projection of these points on the 3D subspace of DCT coefficients with indices $i=5,8,11$. The left panel of Fig. 8(b) shows the time course of this detector's sensor data as it ''heard'' the sound in Fig. 5(a), and the right panel shows the rescaled sensor state representation. Comparison of Figs. 5(c) and 8(b) shows that the two simulated detectors produced nearly identical rescaled sensor-state representations, despite the fact that they were ''equipped'' with different sensors. This is expected because the sensor states of the two detectors were related by an invertible mapping. To see this, recall that each detector was sensitive to the differences among the sound's spectra. Therefore, in each detector the sensor state induced by a spectrum was invertibly related to the position of that spectrum within the one-dimensional collection of the sound's spectra in the space of all spectral parameters.

One can argue that rescaled representations are insensi-

tive to the way sound spectra are filtered or parametrized because they reflect an ''inner'' property of the motion of the warbler's vocal tract, rather than reflecting the ''outer'' signal propagation and detection processes. Because the spectra of the warbler's song were confined to a one-dimensional trajectory in the space of all possible spectra, it is likely that the warbler's vocal tract moved through a one-dimensional manifold in the space of all possible configurations (e.g., its vocal-tract muscles were continuously reconfigured in a coordinated fashion that depended on just one time-dependent parameter). If the sensed DCT coefficients were sensitive to each of these motions, there must be an invertible mapping between the detector's sensor state (x) and the inner parameter that describes the configuration of the vocal tract. It follows that the time courses of the sensor state and this parameter must have identical rescaled representations; i.e., the sound's rescaled representation describes an intrinsic property of the vocal tract's motion that is independent of the way it is observed. The next section explicitly demonstrates this phenomenon in the context of a family of simulated vocal tracts with one degree of freedom.

V. EXPERIMENTS WITH SPECTRA OF SYNTHETIC SPEECH-LIKE SOUNDS

In this section, we consider a family of simple simulated vocal tracts that are configured by varying a single parameter, and, in the context of these models, we demonstrate the ''speaker'' independence of the rescaled sound representation. Specifically, we consider a pair of sounds that are generated by two different simulated vocal tracts controlled by the same time-dependent parameter, and we show that these sounds have identical rescaled representations.

Each sound was generated by a standard linear prediction (LP) model.[16] In other words, the signals' short-term Fourier spectra were equal to the product of an ''all-pole'' transfer function and a ''glottal'' excitation function. The transfer function had six poles, two real and four complex (forming two complex conjugate pairs). The resulting sound spectra depended on the values of eight real quantities, six that described the positions of the poles and two that described the pitch and amplitude (''gain'') of the excitation. Each of these quantities was a function of a single parameter (g), which itself depended on time. These eight functions described the physical nature of the simulated vocal tract, in the sense that they defined the manifold of all spectra that it could produce as g ranged over all of its possible values. The actual sound produced at any given time was determined by these eight functions, together with the value of $g(t)$. The latter function defined the ''articulatory gesture'' associated with the sound, in the sense that it determined how the simulated vocal apparatus was configured at each time. In a musical analogy, the g-dependent functions of the LP model would describe the possible spectra produced by a musical instrument played with one finger, and the function $g(t)$ would describe the motions of the musician's finger as it configures the instrument during a particular tune. In these examples, we considered ''voiced'' sounds that were driven by regular excitation functions. However, it is straightforward to apply the same methods to ''unvoiced'' sounds that are driven by noise-like excitation functions. In this experiment, we considered two simulated vocal tracts that were described by LP models, whose transfer functions were markedly different functions of g. The two vocal tracts were driven by regular excitation functions with pitches of 200 and 125 Hz, respectively.

Figures 9(b) and 10(a) show the spectrograms of sounds that were generated when each of these simulated vocal tracts was controlled by the ''gesture'' function in Fig. 9(a). These spectrograms were created by digitizing each sound at 10 kHz and then performing short-term Fourier transformations in 10-ms Hamming windows that were advanced in increments of 5 ms. Although the sounds are audibly different, there is a noticeable similarity in the underlying pattern of variation (see the sounds posted at the web site in Ref. 15).

Each sound was sensed by a detector that was simulated as follows. The spectrum at each time point was parametrized by the discrete cosine transformation (DCT) of its magnitude after it had been averaged in equally spaced 600-Hz bins. The simulated detector was assumed to sense the three coefficients with indices $i=2,3,6$, and each sound spectrum defined a single point in the corresponding 3D space. These points fell on a curve generated by the DCT coefficients of the spectra produced by all possible configurations of the simulated vocal tract (i.e., all possible values of g). Figures 9(c) and 10(b) show the different configurations of this curve for the above-described vocal-tract models. The precise shape of each curve depended on the nature of the modeled vocal tract (i.e., on the nature of the g dependence of the model's poles and other parameters). As in Sec. IV, the detector's sensor state consisted of the position of the detected spectrum as it moved along the curve traversed by the sound. This position was measured in a coordinate system (denoted by x) that was established on each curve by projecting each of its points onto a connected array of chords that hugged the curve. As in Sec. IV, each x coordinate system was defined so that its origin coincided with the position of the first spectrum in the sound. The left panels of Figs. 9(d) and 10(c) show the sensory signals produced by the sounds from the first and second simulated vocal tracts [Figs. 9(b) and 10(a)], and the right panels show their corresponding rescaled representations ($\Delta T=500$ ms). Notice that the rescaled representations are nearly identical despite the differences between the simulated vocal tracts, sounds, and sensor states that produced them. This is because the two sensor signals were related by an invertible transformation that preserved the null state. This follows from the fact that the quantity g parametrizes the curves in Figs. 9(c) and 10(b), and, therefore, there is an invertible transformation between $g(t)$ and the sensor signal $x(t)$ for each sound. For the same reason, the rescaled representation of $g_1(t)\equiv g(t)-g(0)$ is identical to the rescaled representations of $x(t)$ for each sound. This means that the rescaled representation of each sound is an inner property of the vocal-tract motion that produced it (i.e., a property that is independent of how the sound propagates or is detected).

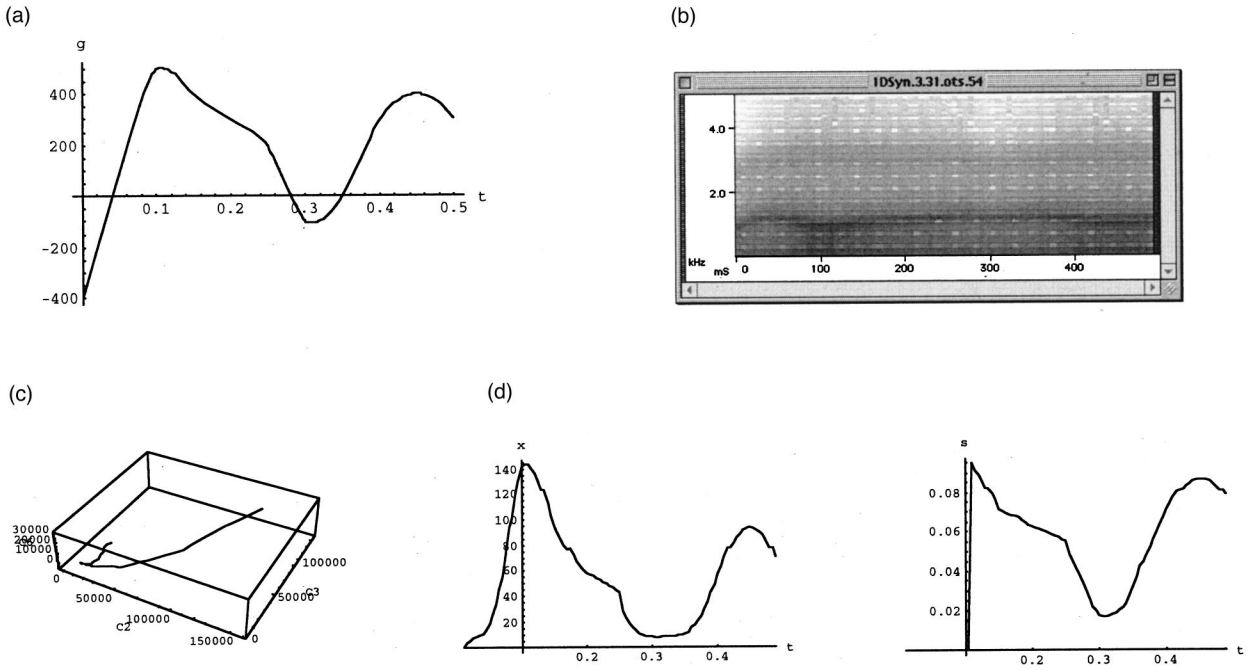


FIG. 9. Simulated vocal tract #1. (a) The time course of the parameter g that described how each simulated vocal apparatus was configured as it generated a sound. Time is in seconds. (b) The spectrogram of the sound produced by the first vocal tract while it was controlled by the ''articulatory gesture'' shown in (a). Time is in ms. (c) The curve swept out by the second, third, and sixth DCT coefficients of the spectra produced by the first simulated vocal tract when it passed through all of its possible configurations (i.e., when the parameter g passed through all of its possible values). (d) Left panel: the sensor state of a detector that measured the position of each sound spectrum in (b) along the curve in (c). Right: the rescaled representation of the sensor data on the left.

VI. DISCUSSION

This paper describes a nonlinear signal processing technique for identifying the ''part'' of a signal that is invariant under any invertible signal distortion. This form of the signal is found by rescaling the signal at each time, in a manner that is determined by its time course in a chosen time interval. The rescaled signal (called its s representation) is unchanged if the original signal time series is subjected to any time-independent invertible transformation. Therefore, if a transmitter encodes information in this representation, it will be faithfully communicated to the receiver despite severe distortions of the propagating signal. This technique can also be used to establish communication among systems with heterogeneous transmitters and receivers that differ by unknown invertible mappings. Such a communication system resembles speech in the sense that: (1) the same information is carried by signals that are related to one another by a wide variety of distortions; (2) the transmitter and receiver can be ''blind'' to the nature of the distortion, which remains unknown; (3) if the distortion changes, faithful communication resumes after a period of adaptation.

The method in this paper differs significantly from techniques for multidimensional scaling or dimensional reduction.[17–19] In each of these methods, it is necessary to impose an *ad hoc* measure of ''distance'' between each pair of neighboring data points[20] or, at least, to rank the distances between pairs of neighboring points.[21,22] In each case, the defined distances or rankings are not invariant under general, nonlinear coordinate transformations. Therefore, the scale values assigned to each data point are also not transformation independent, unlike the rescaled representations described in this paper. However, it should also be mentioned that multidimensional scaling methods are applicable to data that do not form a time series, unlike the technique in this paper.

Signals have the same s representation, as long as they are related to one another by *time-independent* invertible transformations. Signals that are related by a time-dependent distortion may have different s representations immediately after each change in the nature of the distortion (e.g., Fig. 3). However, the rescaling process eventually adapts to the new form of the distortion, and the invariance of the signal's s representation is re-established. The length of this period of adaptation is ΔT, the user-defined parameter that determines the length of the signal history that is used to derive the scale at each time point. Decreasing ΔT can reduce the duration of this transient corruption of the s representation. However, this strategy will tend to reduce the *number* of signal time derivatives contributing to the computation of the signal scale, and thereby increase the noise sensitivity of the scaling process. Conversely, the noise sensitivity of the s representation can be limited by increasingΔT, at the cost of increasing the time required for the rescaling process to adapt to changes in signal distortion.

The family of all signals $x(t)$ that rescale to a given function $s(t)$ can be considered to form an equivalence class. If such a class includes a given signal, it also includes all invertible transformations of that signal. Signals can be assigned to even larger equivalence classes of all signals that lead to the same result when rescaling is applied N times in succession, where $N \geqslant 2$. For example, suppose that two signals do not have the same representation after one application of rescaling, but the same function is produced by two applications of rescaling. Then, the two signals can be considered to be equivalent at a deeper level, in the sense that

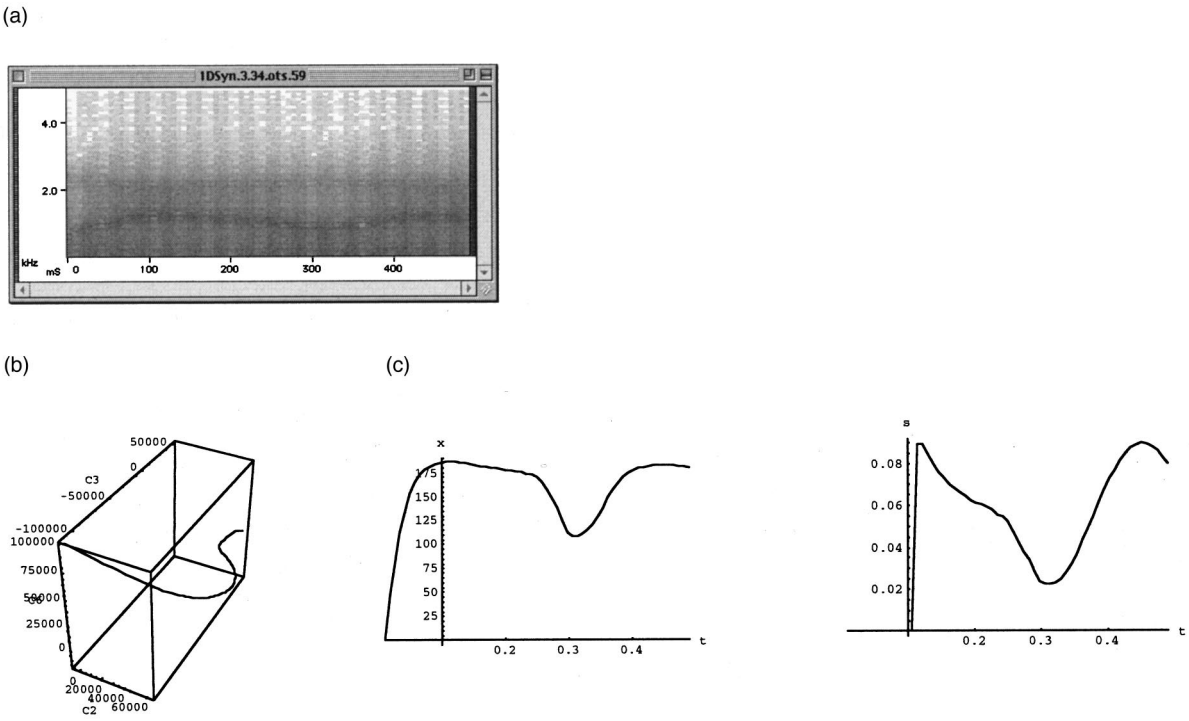


FIG. 10. Simulated vocal tract #2. (a) The spectrogram of the sound produced by the second simulated vocal tract while it was controlled by the ''articulatory gesture'' shown in Fig. 9(a). Time is in ms. (b) The curve swept out by the second, third, and sixth DCT coefficients of the spectra produced by the second simulated vocal tract when it passed through all of its possible configurations (i.e., when the parameter g passed through all of its possible values). (c) Left panel: the sensor state of a detector that measured the position of each sound spectrum in (a) along the curve in (b). Right: the rescaled representation of the sensor data on the left.

the inner forms of their inner forms are the same. Such a relationship is analogous to the relationship between two spoken sentences or phrases, which have different meanings at a superficial level but have the same internal structure at a deeper level. Notice that successive applications of rescaling may result in more and more time points at which an intrinsic scale cannot be computed [e.g., because $h(y)$ cannot be computed or because it vanishes for some values of y]. In other words, as in Fig. 5(c), the serial rescaling process may segment the signal into information-bearing fragments that are separated by time intervals containing no invariants (analogous to the words or phrases of spoken language). Successive applications of rescaling may eventually create a function that is not changed by further applications of the procedure (i.e., the serial rescaling process may reach a fixed ''point''). For example, it is easy to show that, if the natural scale of a signal is time independent [i.e., if $h(y)$ and $s(x)$ are time independent], it will rescale to such a fixed point. Such a fixed point will also be reached if a signal is rescaled by means of a scale that has a fixed and time-independent form dictated by convention (e.g., if the frequency content of a simple melody is rescaled in terms of the equally tempered scale of Western music).

This technology may be used for encoding information in a speech-like telecommunications system, and it may provide a useful ''front end'' for intelligent sensory devices, such as computer vision and speech recognition systems.[23] The signals from the system's detectors would be rescaled before they are passed to the system's pattern recognition module for higher-level analysis. The s representation of a stimulus is invariant under changes in observational conditions that cause invertible transformations of the states of the system's detectors. Such changes may include: (1) alterations of the internal characteristics of the device's detector (e.g., alterations of a microphone's gain characteristics); (2) changes in the observational environment that is external to the sensory device and the stimuli (e.g., changes in the acoustic channel); (3) modifications of the presentation of the stimuli themselves (e.g., modifications of the speaker's voice). Unlike conventional sensory systems, a device with this type of ''representation engine'' need not be periodically recalibrated with test stimuli, and its pattern recognition software need not be retrained when conditions change. This is advantageous because calibration procedures may be logistically impractical in some situations (remote, unsupervised devices), and, in any event, they reduce the device's duty cycle by taking it ''off-line.''

In Sec. IV, rescaling was demonstrated by applying it to a bird song whose spectra were filtered and parametrized in a variety of ways. The results suggested that the rescaled representation of a sound is insensitive to the exact nature of the filtering and parameterization operations, as long as these procedures do not obscure the differences between the various spectra contained in the sound. In Sec. V, sounds produced by different vocal-tract models had the same rescaled representations, as long as there was an invertible mapping between the underlying parameters controlling the vocal-tract configurations. One can argue that rescaled representations will exhibit this kind of invariance for even more general classes of vocal tracts and detectors than those

exemplified in Secs. IV and V. To see this, assume that a detector of interest is sensitive to the differences between the spectra generated by any two configurations of a simulated vocal apparatus. It follows that those configurations are invertibly related to the detector's sensory states when the vocal-tract model is used to create sounds. Therefore, if sound is sensed by two different detectors with this sensitivity, their sensory states will be invertibly related to one another and, consequently, have identical rescaled representations. Similarly, consider a single detector that sensitively records the sounds from two different vocal-tract models. Assume that there is an invertible transformation between the configurations of the two vocal tracts when they generate comparable sounds. This might happen because one vocal tract mimics the other in a consistent fashion (in analogy with two speakers ''reading'' from the same ''text'' in a consistent manner). Then, the sensory signals induced in the detector by the two vocal tracts will be invertibly related. This is because these sensory signals are invertibly related to vocal-tract configurations, which are themselves invertibly related. It follows that the detector will construct identical rescaled representations of the sounds from the two vocal tracts. Thus, rescaled sound representations should be speaker independent and detector independent for a wide class of speaker and detector models. Furthermore, as mentioned in Sec. V, because the vocal apparatus configurations producing a sound are invertibly related to the sensory signals induced by the sound, the trajectory of the underlying vocal-tract parameter $[g(t)]$ controlling that configuration will have the same rescaled representation as the sound itself. If this parameter's time course is regarded as a record of the articulatory gesture of the vocal tract and the rescaled sensor state representation is regarded as the ''percept'' induced by the sound, this is consistent with the ''motor'' theory of speech perception.[24] From this point of view, the rescaled representation of a sound represents an inner property of the vocal tract's motion, and that is *why* it is independent of outer features of the detection process (such as the nature of the detectors and the acoustic channel).

Although the experiments in Secs. IV and V were performed with sounds depending on a single underlying degree of freedom, it is straightforward to generalize the methodology to signals produced by vocal tracts with multiple degrees of freedom. For example, consider the spectra generated by a vocal apparatus with two degrees of freedom. The set of all possible spectra from that apparatus defines a 2D subspace (i.e., a sheet-like surface) in the space of spectral parameters (e.g., DCT coefficients), and the shape and location of that subspace characterize the range of possible sounds which that vocal apparatus can produce. Any particular utterance will be characterized by a trajectory on such a 2D surface, and the form of the unrescaled trajectory will generally be dependent on the nature of the detectors, channel, and speaker. In Refs. 10 and 12–14, the author demonstrated several differential geometric techniques for rescaling such signals with two (or more) degrees of freedom. It may be computationally practical to apply these techniques to human speech,[14] because there is evidence that the human vocal tract has a relatively small number of degrees of freedom (e.g., 3–5; see Refs. 25–26). As in the previous paragraph, it can be argued that the same rescaled representation would be generated from any given utterance produced by a wide variety of speakers and propagated through a wide variety of channels and detectors. This is because the different spectral trajectories corresponding to a given utterance can be considered to represent the same trajectory in different coordinate systems, and the rescaling process extracts the coordinate-independent relationship of that trajectory to the intrinsic geometric structure of the speech manifold. Therefore, a speech recognition device with such a front end may not need extensive retraining when the speaker's voice, the acoustic channel, and/or the detectors are changed.

One can also speculate that the adaptive nature of the rescaling process might enable it to account for coarticulation[16] during human speech. Recall that the manner in which each sound (i.e., each parametrized spectrum) is rescaled depends on the nature of recently encountered sounds. It could also depend on the nature of sounds to be encountered in the near future, if the interval ΔT is defined to include times *after* the sound to be rescaled. In other words, the rescaled representation of each sound spectrum depends on its acoustic *context* (defined by the endpoints of ΔT), similar to the contextual dependence of speech perception that is the hallmark of the coarticulation phenomenon. Finally, the foregoing considerations make it tempting to speculate that the human brain itself decodes speech signals by constructing some type of dynamically rescaled version of speech spectra. This could account in part for the ease of speech communication involving a variety of speakers, listeners, and acoustic environments. Of course, the considerations in this paragraph are purely theoretical and must be tested by application to actual human speech spectra.

[1] *Computational Models of Speech Pattern Processing*, edited by K. M. Ponting (Springer, Berlin, 1999).

[2] L. C. Nygaard, and D. B. Pisoni, ''Talker-specific learning in speech perception,'' Percept. Psychophys. **60**, 355–376 (1998).

[3] D. B. Pisoni, ''Some thoughts on normalization in speech perception,'' in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego, 1997), pp. 9–32.

[4] G. M. Stratton, ''Some preliminary experiments on vision without inversion of the retinal image,'' Psychol. Rev. **3**, 611–617 (1896).

[5] G. M. Stratton, ''Vision without inversion of the retinal image,'' Psychol. Rev. **4**, 341–360 (1897).

[6] G. M. Stratton, ''Vision without inversion of the retinal image (concluded),'' Psychol. Rev. **4**, 463–481 (1897).

[7] J. J. Gibson, ''Adaptation, after-effect, and contrast in the perception of curved lines,'' J. Exp. Psychol. **16**, 1–31 (1933).

[8] R. Held and R. Whitman, *Perception: Mechanisms and Models* (Freeman, San Francisco, 1972).

[9] D. N. Levin, ''Time-dependent signal representations that are independent of sensor calibration,'' J. Acoust. Soc. Am. **108**, 2575 (2000).

[10] D. N. Levin, ''Stimulus representations that are invariant under invertible transformations of sensor data,'' in Proceedings of the Society of Photoelectronic Instrumentation Engineers, 4322, pp. 1677–1688 (2001). This paper can be downloaded from http://www.geocities.com/dlevin2001/reprint1.html

[11] D. N. Levin, ''Universal communication among systems with heterogeneous voices and ears,'' in Proceedings of the International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet, Scuola Superiore G. Reiss Romoli S.p.A., L'Aquila, Italy, 6–12 August 2001. This paper can be downloaded from http://www.ssgrr.it/en/ssgrr2001/index.htm or from http://www.geocities.com/dlevin2001/reprint2.html

[12] D. N. Levin, ''Sensor-independent stimulus representations,'' Proc. Nat'l. Acad. Sci. (USA), in press (2002). To download, see http://www.geocities.com/dlevin2001/preprintz.html
[13] D. N. Levin, ''Stimulus representations that are independent of sensors and observational conditions''. To download, see http://www.geocities.com/dlevin2001/preprint3.html
[14] D. N. Levin, ''Blind normalization of speech from different channels and speakers.'' To dowload, see http://www.geocities.com/dlevin2001/reprint3.html
[15] The sounds corresponding to each figure in Secs. III–V can be played or downloaded at http://www.geocities.com/dlevin2001/preprint1.html
[16] L. Rabiner and B-H. Juang, *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ, 1993).
[17] R. N. Shepard, ''The analysis of proximities: Multidimensional scaling with an unknown distance function. I,'' Psychometrika **27**, 125–140 (1962), and II, **27**, 219–246 (1962).
[18] J. D. Carroll and P. Arabie, ''Multidimensional scaling,'' Annu. Rev. Psychol. **31**, 607–649 (1980).
[19] T. Cox and M. Cox, *Multidimensional Scaling* (Chapman & Hall, London, 1994).
[20] S. T. Roweis and L. K. Saul, ''Nonlinear dimensionality reduction by locally linear embedding,'' Science **290**, 2323–2326 (2000).
[21] E. W. Holman, ''Completely nonmetric multidimensional scaling,'' J. Math. Psychol. **18**, 39–51 (1978).
[22] J. B. Tenenbaum, V. de Silva, and J. C. Langford, ''A global geometric framework for nonlinear dimensionality reduction,'' Science **290**, 2319–2323 (2000).
[23] D. N. Levin, Patents pending.
[24] A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, ''Perception of the speech code,'' Psychol. Rev. **74**, 431–461 (1967).
[25] N. Tishby, ''A dynamical systems approach to speech processing,'' in Proceedings, 1990 International Conference on Acoustics, Speech, and Signal Processing, pp. 365–368 (1990).
[26] B. Townshend ''Nonlinear prediction of speech signals,'' *Nonlinear Modeling and Forecasting. Proceedings of the Workshop on Nonlinear Modeling and Forecasting, Sante Fe, NM, 1990*, edited by M. Casdagli and S. Eubank (Addison-Wesley, New York, 1992).

A reflex resonance model of vocal vibrato

Ingo R. Titze
National Center for Voice and Speech, and Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242 and, Wilbur James Gould Voice Center, Denver Center for the Performing Arts, Denver, Colorado 60204

Brad Story
National Center for Voice and Speech, The University of Iowa, Iowa City, Iowa 52242 and Speech and Hearing Sciences, University of Arizona, Tucson, Arizona 85721

Marshall Smith
National Center for Voice and Speech, The University of Iowa, Iowa City, Iowa 52242 and School of Medicine, University of Utah, Salt Lake City, Utah 84112

Russel Long
National Center for Voice and Speech, The University of Iowa, Iowa City, Iowa 52242 and Wilbur James Gould Voice Center, Denver Center for the Performing Arts, Denver, Colorado 60204

(Received 6 November 2000; revised 30 October 2001; accepted 19 November 2001)

A reflex mechanism with a long latency (>40 ms) is implicated as a plausible cause of vocal vibrato. At least one pair of agonist–antagonist muscles that can change vocal-fold length is needed, such as the cricothyroid muscle paired with the thyroarytenoid muscle, or the cricothyroid muscle paired with the lateral cricoarytenoid muscle or a strap muscle. Such an agonist–antagonist muscle pair can produce negative feedback instability in vocal-fold length with this long reflex latency, producing oscillations on the order of 5–7 Hz. It is shown that singers appear to increase the gain in the reflex loop to cultivate the vibrato, which grows out of a spectrum of 0–15-Hz physiologic tremors in raw form. [DOI: 10.1121/1.1434945]

PACS numbers: 43.75.Bc, 43.75.Rs, 43.70.Aj, 43.70.Bk [RDA]

I. INTRODUCTION

Vibrato is an ornament of music that is used by both vocalists and instrumentalists. As a frequency modulation of the fundamental frequency, vibrato enriches the average spectral content of a sustained tone because the harmonics can sweep through acoustic resonances of the instrument. If executed correctly, vibrato draws no attention to itself, but rather enhances the quality of the sound. Listeners generally judge tones with vibrato as being more lively, freer from tension, and more filling of the performance space. But, excessive vibrato creates the perception of instability, lack of clarity of pitch, and a general ''muddiness'' of the melodic line. Skilled performers know the bounds of vibrato, decreasing its extent during rapid melodic movement and increasing its extent during long, sustained passages when melody and vibrato are less likely to be confused.

The classic studies by Seashore and his co-workers (summarized by Seashore, 1938) are still cited often as key authoritative measurements of the acoustic characteristics of vocal vibrato. Seashore wrote, ''A good vibrato is a pulsation of pitch, usually accompanied with synchronous pulsations of loudness and timbre, of such extent and rate as to give a pleasing flexibility, tenderness, and richness to the tone'' (p. 33).

Seashore showed examples of the acoustic characteristics of vocal and violin vibrato. He demonstrated that vocal vibrato *extent* varied significantly from note to note, with the average peak-to-peak variation being one semitone (or $\pm$one quarter tone from the mean F_o). Seashore also measured the vibrato *frequency*. He reported an average of 6.5 Hz, with less note-to-note variability in this frequency than in the extent.

Not much has changed in the acoustic description of vocal vibrato over the last two-thirds of a century. However, vibrato frequencies of classically trained singers have recently been reported slightly lower than in the Seashore studies. To compare two notable tenors of the first and second half of the century, Caruso's average vibrato frequency was 6.5 Hz (Dejonkere, 1995), whereas Pavarotti's frequency has been measured at 5.5 Hz (Keidar, Titze, and Timberlake, 1984). Informally, many vocal pedagogues have observed (by listening to old and new recordings of identical pieces) that vibrato frequency has lowered from the pre-World War II era to the end of the 20th century. There are no compelling explanations, but perhaps a preference toward a more ''settled'' sound (less overt expressions of emotion) is a part of the trend.

In Prame's (1994) study, the average frequency of ten classically trained recording artists singing Schubert's *Ave Maria* was 6.0 Hz (not the same as the Bach–Gounod *Ave Maria* studied by Seashore). Prame noted that all of the singers tended to raise their vibrato frequencies at the end of each note. In particular, the last four to five vibrato cycles typically rose from 6.0 to 7.0 Hz. In the more central portion of the phrase, however, the frequency was often below 6.0 Hz, typically about 5.5 to 5.8 Hz.

Physiologically, vibrato has been described as a sympathetic (if not synergistic) oscillation phenomenon among laryngeal muscles. Hirano (1995) pointed out that the crico-

thryoid (CT) muscle is always activated in an oscillatory fashion in vibrato, whereas the thyroarytenoid (TA) muscle and the lateral cricoarytenoid (LCA) muscles are usually (but not always) involved. It is possible that the CT muscle may even interact with a strap muscle (e.g., the sternothyroid or the sternohyoid, which are known to lower the larynx, and thereby perhaps shorten the vocal-folds). Hirano also reported that there can be oscillatory movement of many surrounding structures, such as the pharyngeal walls, the velum, the tongue, the epiglottis, and the jaw. A particularly interesting aspect of his study was the declining vibrato frequency with age, which he reported to shift from 5.4 Hz at age 20 to 4.7 Hz at age 65, on average, for a group of 21 singers. This decline of about 1 Hz over half a lifespan agreed with earlier findings by Damsté *et al.* (1982) and later findings by Sundberg *et al.* (1998). Both groups of authors measured the vibrato frequencies of several professional singers who recorded the same pieces several times over a span of 30–40 years.

Summarizing the preliminary findings on the physiology of vocal vibrato in our own laboratory, Titze *et al.* (1994) observed a 5–6-Hz reflex resonance mechanism for vocal vibrato that involved the CT and TA muscles. The study was conducted only on one singer, an amateur tenor. Damped oscillation was postulated to result from a negative feedback loop of moderate gain, the nature of which will be explored in greater detail here. When combined with the mechanics of rotation and translation of the thyroid and cricoid cartilage relative to each other, Titze (1996) showed that oscillatory behavior in vocal-fold length, tension, and fundamental frequency could be simulated with a mathematical version of this reflex oscillator. Hsiao *et al.* (1988) showed that motor unit activity in the CT muscle was correlated with F_o changes in vibrato, both for integrated electromyographic (EMG) activity and single motor unit activity. Such correlations were also evident in laryngeal tremors studied by Finnegan *et al.* (1995), but the phase relations between CT, TA, and voice F_o remained somewhat of a mystery.

A complete physiologic model of vocal vibrato may eventually explain the following empirical observations, mostly untested but somewhat agreed upon by singing teachers: (1) vibrato frequency is less controllable by the singer than vibrato extent; (2) many children (and some adults) sing vibrato-free; (3) many elderly or poorly conditioned vocalists have a slower and wider vibrato (a wobble), which can be corrected by singing soft and high and by strengthening laryngeal muscles (personal conversations with singing teachers); (4) some young or muscularly hyperactive vocalists have an excessively fast vibrato (a bleat) that can be eliminated by singing in the midrange, using muscle relaxation exercises, and establishing a balanced vocal quality (personal experience gained by the first author in training singers); (5) training in general reduces the undesirable variability in vibrato frequency; (6) violin, brass, or woodwind vibrato can be of similar rate as vocal vibrato, suggesting that the mass of the undulating structure (e.g., the hand or respiratory bulk) doesn't dictate the rate as much as neurologic factors, as noted by Ramig and Shipp (1987); (7) some pathological vocal tremors are close to natural vibrato (5–6 Hz); (8) vibrato rate often increases at the end of a note, as reported by Prame (1994); (9) vibrato occurs more naturally in the midrange of F_o than at the extremes (very low F_o or very high F_o); and (10) vibrato rate can be entrained by an external stimulus (Titze *et al.*, 1994). In this paper, we will address a few of these observations, but the investigation leaves many questions unanswered.

II. THE REFLEX RESONANCE MODEL

The proposed reflex resonance model for vibrato is schematized in Fig. 1. On the far right of the figure, the physiological processes are divided, from top down, into cortical activity, midbrain activity, and peripheral activity. The input to the model (at the top) is the pitch intended by the singer, with a vibrant quality but no deliberate attempt to create a vibrato. The output (at the bottom) is the fundamental frequency F_o, which has a frequency modulation ranging between 4–7 Hz, but the half-power bandwidth is more like 5–6 Hz.

Cortically (top third of the diagram), the motor program gets activated by auditory and kinesthetic memory of what a vibrant tone at a given pitch sounds and feels like, as well as the muscular effort and coordination it takes to produce it. In the process of organizing the muscle activations, central oscillators are activated that superimpose a broad spectrum of physiologic tremor freqencies onto an otherwise constant level of activation (Inbar and Eden, 1983; Elble and Koller, 1990). The slightly tremorous motor activations to the CT and TA muscles are labeled A_{CT} and A_{TA}, respectively. The model is not limited to these two muscles. Other agonist–antagonist pairs (e.g., a strap muscle working together with either the TA or CT, or the LCA working together with the CT) could be engaged in this reflex resonance pairing, as long as they have an opposite effect on vocal-fold length and tension (and thereby F_o).

In the midbrain (center of the diagram), there is an integration of sensory signals arriving from the laryngeal receptor organs (muscle spindles, joint receptors, etc.) and the cortical motor activations. The sensory-motor integration is symbolized by the summation signs and involves primarily corrective feedback in the form of time-delayed muscle strains $\varepsilon_1(t-\tau_1)$ and $\varepsilon_2(t-\tau_2)$, multiplied by their respective gains g_1 and g_2. The feedback-modulated activations are labeled a_{CT} and a_{TA}, respectively. The important modeling parameters here are g_1 and g_2, the feedback gains, and τ_1 and τ_2, the feedback delays. These parameters will be given considerable attention in later discussions of experiments and modeling.

At the periphery, the CT and TA muscles are now activated with the modulated action potentials. This in turn modulates the vocal-fold length, which is quantified in terms of the vocal-fold strain ε, the elongation normalized to the resting length. If vocal-fold strain increases, the muscle spindles in the thyrovocalic portion of the TA muscle are activated (Sanders *et al.*, 1998). With the appropriate gain g_2 and the reflex time delay τ_2, this becomes the sensory modulation on the right side of Fig. 1 (dashed lines). If the vocal-fold strain decreases, the muscle spindles in the TA are not activated, but joint receptors or spindles in the CT muscle

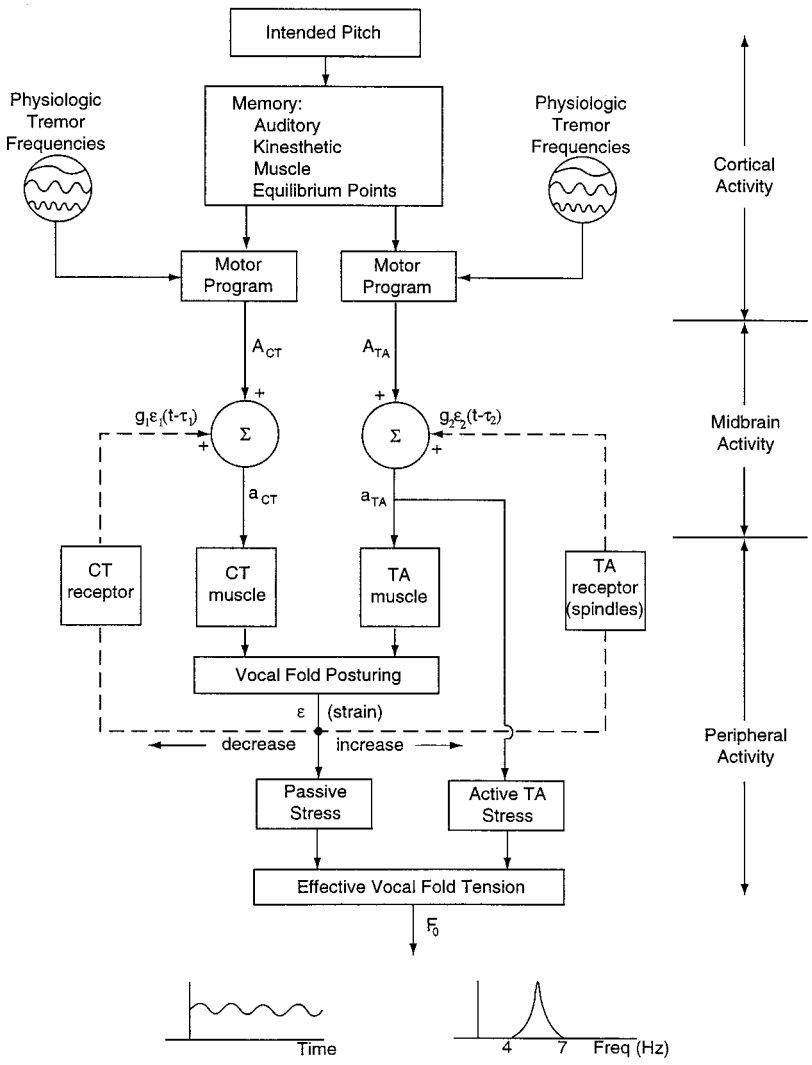


FIG. 1. Block diagram of proposed reflex-resonance model of vocal vibrato.

(which is now stretched) are activated. In this case, the CT muscle strain ε_1 is not equal to the vocal-fold strain ε, but is a predictable function of it (Titze, 1996). With respective gain g_2, this feedback becomes the sensory reflex modulation for the CT muscle on the left side of the figure, again shown in dashed lines.

Concurrently, the vocal-fold strain ε determines the passive tissue fiber stress in the vocal-fold, and the activation a_{TA} determines the active stress in the muscular portion of the fold (bottom three system blocks in Fig. 1). Together, the two stresses govern the effective tension, and thereby the fundamental frequency F_o. The basic hypothesis in this model is that the reflex gains and delays resonate, thereby tuning the broad spectrum of central tremor frequencies to a narrower band, which is known as vibrato (illustrated in temporal and spectral form at the bottom of Fig. 1). Thus, singers with vibrato are expected to have a narrower bandwidth (with correspondingly higher gains) in the reflex loops than those who phonate without vibrato.

The time course of events is hypothesized to be as in Fig. 2. Assume we begin with an increase in CT activation, represented by a small spike disturbance in Fig. 2(a) at $t=0$. (For simplicity, we illustrate all neural signals as single spikes, even though they are typically trains of spikes, within and across motor and sensory units. The single spike is considered to represent the center of any train of action potentials.) Below the spike shown in Fig. 2(a), a twitch-like contraction occurs in the CT muscle [Fig. 2(b)], which reaches its peak about 33 ms after activation (Perlman and Alipour-Haghighi, 1988). The vocal-fold length (strain) increases in proportion to this contraction, with a small mechanical delay (10–20 ms), as shown in Fig. 2(c). This length increase triggers the TA muscle spindles, which are sensitive to both strain ε and strain rate $\dot{\varepsilon}$ (Stein, 1982). Spindle activation is shown in Fig. 2(d). The strain-rate dependence advances the phase, in part canceling out the mechanical delay in vocal-fold elongation, as observed by Stein.

The major time delay now comes from the integration of the spindle sensory activation with the motor activation, which is labeled TA reflex motor activation in Fig. 2(e). The time delay is on the order of 40–50 ms, as we will show. A TA contraction now occurs, the peak of which occurs another 20–25 ms after the activation (Alipour-Haghighi *et al.*, 1987). This is shown in Fig. 2(f).

Vocal-fold strain ε is now reduced [Fig. 2(g)], with a concomitant increase in the length of the CT muscle and a receptor activation [Fig. 2(h)]. The reflex loop is engaged once more, resulting in another delay for sensory-motor in-

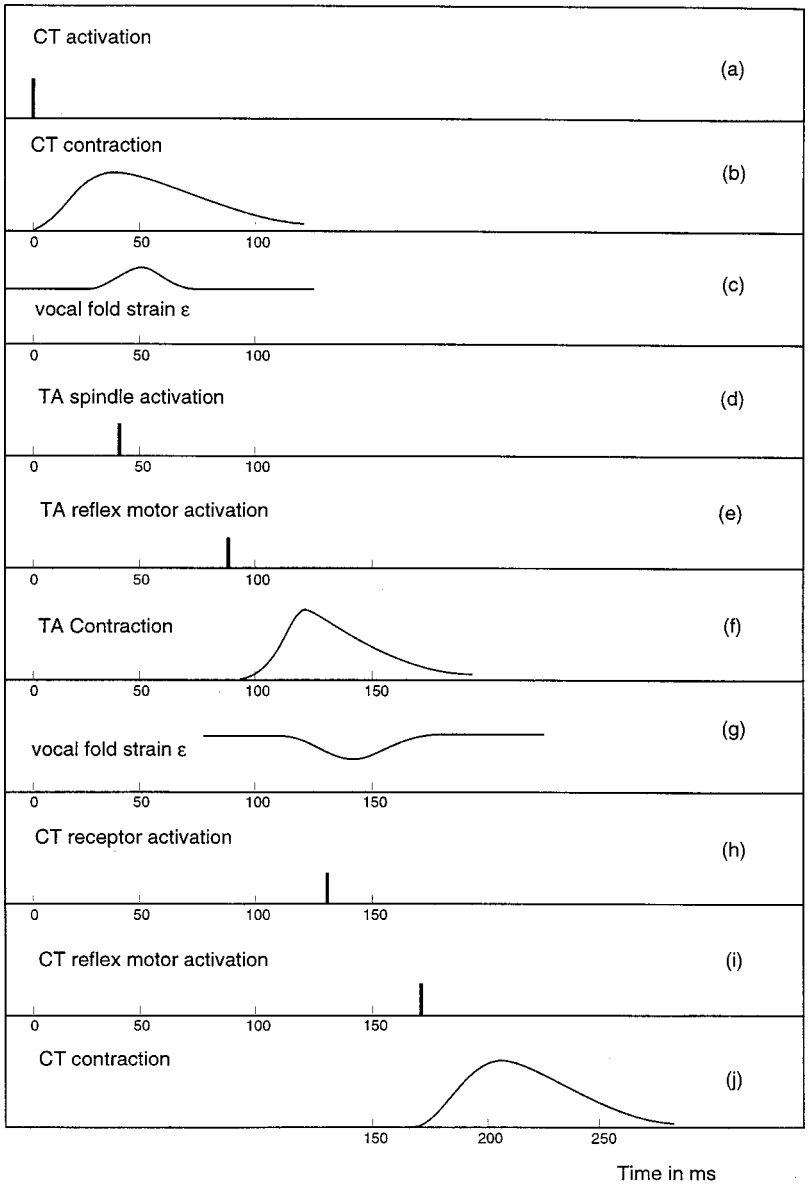


FIG. 2. Time course of efferent and afferent stimuli, and muscle activation.

tegration [Fig. 2(i)] and another CT contraction [Fig. 2(j)]. The complete cycle, from the peak of CT contraction to the peak of CT contraction, has a period of about 170 ms, yielding a vibrato frequency of about 6 Hz.

III. DYNAMICS OF THE MUSCLES AND OTHER SOFT TISSUES

Mathematically, the model begins by letting both muscles be represented mechanically by a Kelvin model (McMahon, 1984, p. 136), as shown in Fig. 3. This model has a series spring k_s, a parallel spring k_p, a damper d, and an internal contractile stress σ_i (in the case of muscle). The onset of the internal contractile stress is time delayed by a first-order differential equation (Stein and Oguztöreli, 1975)

$$\dot{\sigma}_i=(\sigma_a-\sigma_i)/t_a \quad i=1,2, \tag{1}$$

where the ''dot-over'' notation indicates time differentiation, t_a is the activation time constant, and σ_a is the fully developed active stress.

A second differential equation for the muscle model is obtained by writing Newton's law for the internal (left) nodal point indicated in Fig. 3

$$S\sigma_i-k_sL_o(\varepsilon_i+\varepsilon)-dL_o\dot{\varepsilon}_i=0, \tag{2}$$

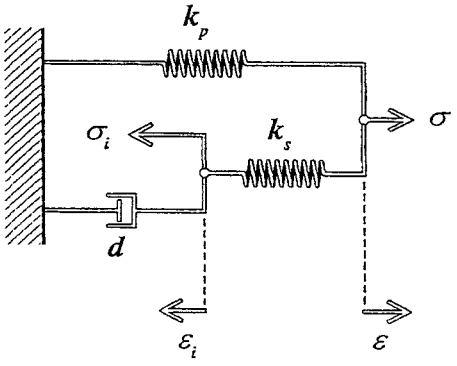


FIG. 3. Kelvin muscle model for twitch and tetanus force computation.

TABLE I. Biomechanical constants of the cricothyroid muscle.

	Symbol	Value
Resting length	L_o	13.8 mm
Cross-sectional area	S	73.8 mm^2
Mass	M	0.9423 g
Maximum active stress	σ_{am}	89 kPa
Activation time	t_a	0.01 s
Series contraction time	t_s	0.090 s

TABLE II. Biomechanical constants of the thyroarytenoid muscle.

	Symbol	Value
Resting length	L_o	18.3 mm
Cross-sectional area	S	40.9 mm^2
Mass	M	0.8232 g
Maximum active stress	σ_{am}	105 kPa
Activation time	t_a	0.01 s
Series contraction time	t_s	0.044 s

and a nondifferential equation is obtained by writing Newton's law for the external (right) nodal point

$$F - k_p L_o \varepsilon - k_s L_o(\varepsilon_i + \varepsilon) = 0, \tag{3}$$

where S is the cross-sectional area of the muscle, L_o is the rest length, ε_i is the internal (sarcomere) strain, ε is the external muscle strain, and F is the external (load) force. Equation (2), the nondifferential equation, is now solved for ε_i and substituted into Eq. (3) to yield

$$S\sigma_i - F + k_p L_o \varepsilon - \frac{d}{k_s}\dot{F} + d\frac{k_p + k_s}{k_s} L_o \dot{\varepsilon} = 0. \tag{4}$$

To simplify the model and capture only its essence in this first exposition of vibrato, the passive tissue force $k_p L_o \varepsilon$ will be neglected. This passive force is usually an order of magnitude smaller than the active force (Alipour and Titze, 1999) except for high strains (30%–50% of rest length). We will keep the strains below 10% in our simulations. Furthermore, the strain rate is low in vibrato, which allows the $\dot{\varepsilon}$ term in Eq. (3) to be neglected also in comparison to the active fiber tension $S\sigma_i$.

With these simplifications, Eq. (3) reduces to

$$\dot{F} = [S\sigma_i - F]/t_s, \tag{5}$$

where

$$t_s = \frac{d}{k_s}. \tag{6}$$

Equations (1) and (4) develop the load force in each muscle, based on activation delay and active stress in the fibers. The series contraction time t_s of the sarcomere is part of the overall delay of response of the muscle system (Stein and Oguztöreli, 1975). Tables I and II show measured values of the parameters for the CT and TA muscles, respectively. These values were obtained in our laboratory over a period of more than a decade (Alipour-Haghighi, Titze, and Durham, 1987; Perlman and Alipour-Haghighi, 1988; Alipour-Haghighi, Titze, and Perlman, 1989; Alipour, Perlman, and Titze, 1991; Alipour and Titze, 1999). The properties were obtained from a canine model, however, and must be regarded as approximations to human tissues.

The fully developed active stress function σ_a in Eq. (1) is generated for each muscle as

$$\sigma_a = a(t)\sigma_{am}, \tag{7}$$

where $a(t)$ is the muscle activation (ranging from 0 to 1) and σ_{am} is the maximum active force achievable in the muscle under isometric conditions (Tables I and II). Usually this maximum active force varies with muscle strain ε, but to keep the model simple we will keep it a constant value, which is justifiable because strain variations are small in vibrato.

IV. DYNAMICS OF THE CARTILAGE FRAMEWORK

The differential equations of motion for translation and rotation of the cricoid cartilage in relation to the thyroid cartilage were developed previously (Titze, 1996). These equations are rewritten here for completeness. For translation ψ and rotation θ around the cricothyroid joint (Fig. 4)

$$\ddot{\psi} = \frac{1}{M}[\cos\phi F_{\mathrm{CT}} - F_{\mathrm{TA}} - k_t(\psi + t_t\dot{\psi})], \tag{8}$$

$$\ddot{\theta} = \frac{1}{I}[wF_{\mathrm{CT}} - hF_{\mathrm{TA}} - k_r(\theta + t_r\dot{\theta})], \tag{9}$$

where F_{CT} is the load force generated by the CT muscle, F_{TA} is the load force generated by the TA muscle, $\cos\phi = 0.76$ is the direction cosine between CT and TA lines of action, $w = 0.011$ m is the moment arm for CT force in CT joint rotation, $h = 0.016$ m is the moment arm for TA force in CT joint rotation, $M = 0.01$ kg is the mass of cricoid and arytenoid cartilages translating as a unit (relative to the thyroid cartilage), $I = 10^{-5}$ kg m^2 = moment of inertia of cricoid and arytenoid cartilages rotating as a unit around the CT joint, k_t is the stiffness of CT joint in translation (defined below), k_r is the torsional stiffness of the CT joint in rotation (defined below), $t_t = 0.010$ s is the time constant for translation, specifying viscous damping, and $t_r = 0.010$ s is the time constant for CT joint rotation, specifying viscous damping.

The strains ε_1 and ε_2 of the CT and TA muscles, respectively, are related to θ and ψ as follows:

$$\varepsilon_1 = -(w\theta + \psi/\cos\phi)/L_{01}, \tag{10}$$

$$\varepsilon_2 = (h\theta + \psi)/L_{02}, \tag{11}$$

where L_{01} and L_{02} are the rest lengths of the muscles. These are the strains used for the receptor pathways in Fig. 1.

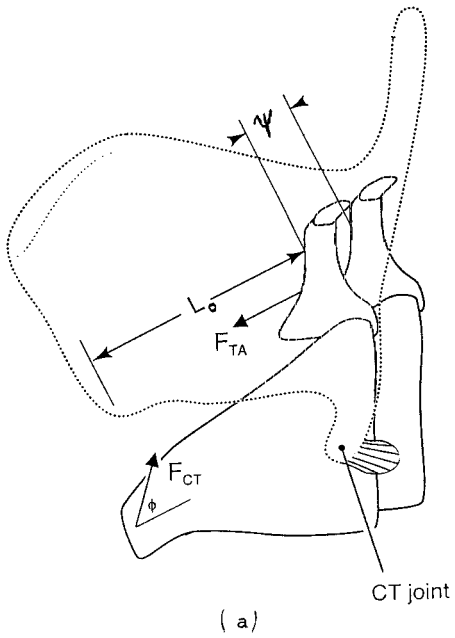
The translational stiffness k_t in Eq. (8) has been measured as a nonlinear function of strain by Vilkman *et al.* (1987). We give the following approximation to their data:

$$k_t = 1000 + 10^5(\psi/L_{01})^2 \text{ N/m}. \tag{12}$$

Similarly, the rotational stiffness has been measured by Cooper *et al.* (1993), approximated here as

$$k_r = 0.1 + 10(h\theta/L_{01})^2 \text{ N-m/radian}. \tag{13}$$

The nonlinear nature of these stiffnesses provides the limit cycle for oscillation in vibrato. If the stiffnesses were linear,



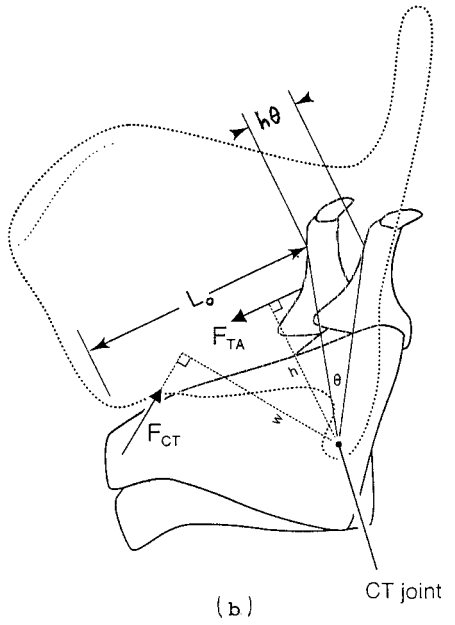


FIG. 4. Laryngeal cartilages in (a) translation and (b) rotation.

underdamped oscillation would continue to grow beyond bound.

V. DYNAMICS OF SENSORY FEEDBACK

Feedback from the sensory organs is modeled by letting the modulated muscle activities be

$$a_{\mathrm{CT}}(t)=A_{\mathrm{CT}}(t)+g_1[\varepsilon_1(t-\tau_1)+0.03\dot{\varepsilon}_1(t-\tau_1)], \quad (14)$$

$$a_{\mathrm{TA}}(t)=A_{\mathrm{TA}}(t)+g_2[\varepsilon_2(t-\tau_2)+0.03\dot{\varepsilon}_2(t-\tau_2)], \quad (15)$$

where all quantities are as previously defined. $A_{\mathrm{CT}}(t)$ and $A_{\mathrm{TA}}(t)$ are taken as a constant (tetanic activation) with superimposed low-frequency noise that simulates physiologic muscle tremor. The coefficient 0.03 is chosen to weight the strain and strain-rate dependence of the delayed feedback equally, since for sinusoidal strain $\dot{\varepsilon}=\omega\varepsilon$ and ω is about 30 rad/s for a 5–6-Hz vibrato rate.

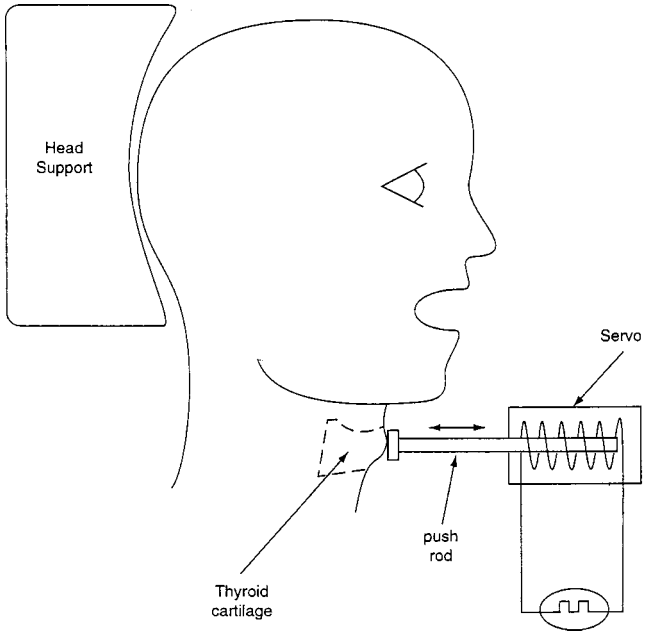


FIG. 5. Experimental setup for mechanical stimulation.

This completes the mathematical description of the reflex resonance model. Parameters that are of primary interest in this paper are the reflex delay times τ_1 and τ_2 and the reflex gains g_1 and g_2.

VI. EXPERIMENTS TO DETERMINE REFLEX GAINS AND DELAYS

Two separate experiments are described in this section. The first experiment involved mechanical stimulation to the larynx to determine the reflex latency. The second experiment involved electromyographic (EMG) recordings to determine the sharpness of the reflex resonance, which we claim is an indirect measure of the reflex gain.

A. Mechanical stimulation

An experiment was conducted to measure the reflex delay associated with mechanical perturbation of vocal-fold length. Two nonsingers, both adult males, served as subjects. EMG electrodes were inserted into both CT and TA muscles to monitor their laryngeal reflexes. A mechanical stimulator was designed in the form of a servo-controlled rod that could push against the thyroid cartilage prominence (Adam's apple) as shown in Fig. 5. The subject sat in a chair in an upright position. His head was secured in place by a headrest. The servo-controlled stimulator was mounted on a rig attached to the chair. The rod had several mm of electronically controllable travel distance after it had been fixed in place, in near-contact with the thyroid cartilage. Thus, the experimenter could adjust the neutral position as ''light contact,'' ''near-contact,'' or ''pressured contact.''

Because the thyroid prominence is rounded, and because the endplate of the rod did not have an ideal cup shape to fit over all the subjects' thyroid prominences, there was occa-

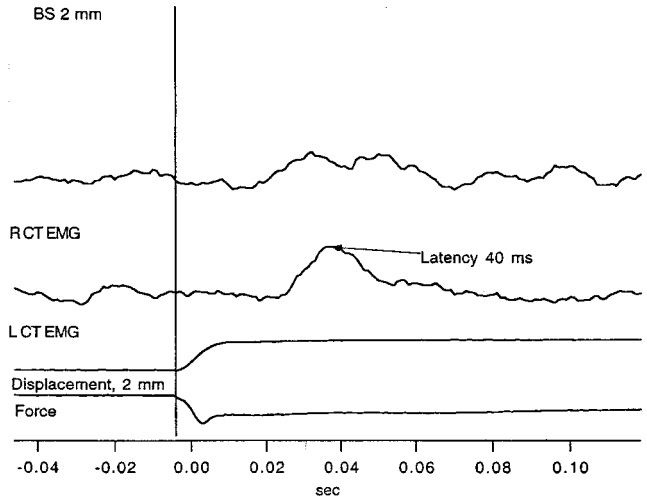


FIG. 6. Reflex latencies for mechanical stimulation.

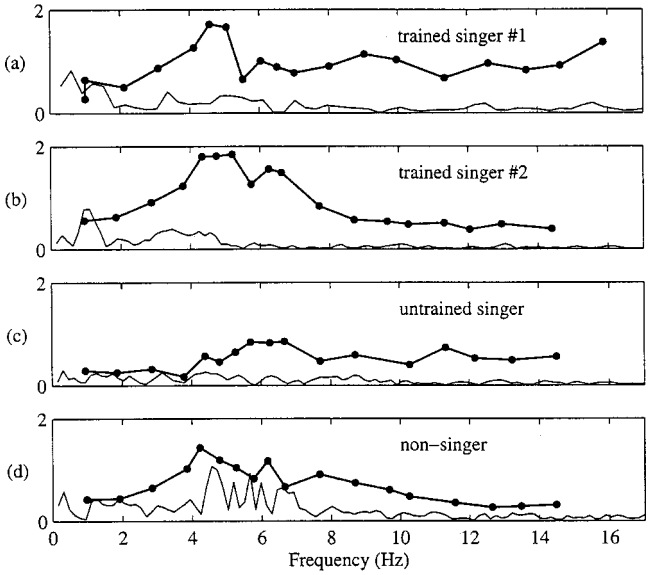


FIG. 7. Resonance curves (amplitude of F_o modulation in Hz) for: CT stimulation and no natural vibrato (data points); no stimulation and no natural vibrato (curves without data points).

sionally some slippage of the rod off-center. This was noted in the experiment, and the data were accepted with caution, or rejected if the slippage was too severe.

Once the neutral position was adjusted, the rod displacement was varied in single-pulse or square-wave fashion with a preset displacement amplitude and frequency (for the square wave). A decrease in vocal-fold strain ε was achieved by pushing against the thyroid cartilage. According to our model, this should activate the CT receptors (spindles or joint receptors) and produce an increase in the motor activation a_{CT} to the CT muscle. The objective was to measure the reflex latency τ_1.

Figure 6 shows results for subject S1. Transducer displacement and force are shown at the bottom (upward displacement means a push toward the cartilage, with an expected negative vocal-fold strain). The displacement amplitude was 2 mm, but there was no knowledge of how much shortening of the vocal-fold this actually produced. The entire larynx was obviously pushed rostrally, which meant that the arytenoid cartilages may have moved together with the thyroid cartilage to produce a lesser change in vocal-fold length. Nevertheless, some reduction in length was expected. The EMG responses of the right CT muscle (R CT EMG) and the left CT muscle (L CT EMG) are shown in the top two traces in Fig. 6. Note that the reflex latency between the maximum velocity of movement (time=0.00) and the peak of the EMG increase is about 40 ms.

Results for the second subject were similar, with the reflex latency being about 44 ms from the maximum movement velocity to the peak of the EMG increase. Both CT muscles showed an increase in activity, but the TA muscles did not. This average reflex latency of about 42 ms for the two subjects was built into our reflex resonance model as the variable τ_1. We had no way of pulling on the cartilage to *elongate* the vocal-folds and obtain a TA reflexive contraction. Hence, τ_2 was chosen to be the same as τ_1 for lack of better data. If the two time constants were in fact different, a slight asymmetry would be expected in the vibrato cycle, and the overall vibrato rate would increase or decrease slightly.

B. Electrical stimulation

Electrical muscle stimulation experiments were conducted on four new subjects, one soprano trained in classical and theatre styles (more than 10 years of private training), one mezzo-soprano trained in classical music (more than 5 years of private training), one untrained female pop singer (no private lessons), and one female nonsinger. (New subjects were used because repeated electrode insertion in close time intervals is not recommended for a given subject, especially not for singers.) Electrical stimulation was applied to the CT muscle during phonation with bipolar hooked-wire electrodes, the same electrodes normally used to record EMG signals (Titze *et al.*, 1994). The resonance condition of the reflex loops was investigated by repeated pulse stimulation of the CT at different frequencies during phonation. According to our model, this would involve the TA reflex as well as the CT reflex. Those subjects who had a natural vibrato were asked to sustain the vowels vibrato-free. Our hypothesis was that if the same pulse current and duration could produce a larger F_o swing at one frequency than all others, then a resonance of the reflex system must exist. Figure 7 shows the results in the form of amplitude spectra of the F_o contour for the four female subjects. The curves without data points show the spectrum of physiologic tremor without any stimulation, whereas the curves with data points are for stimulation at the frequencies shown by the points. With stimulation, we see a tuning curve with a 5-Hz resonance frequency for the two trained singers [parts (a) and (b)] and about 6 Hz for the untrained singer [part (c)]. For the one nonsinger [part (d)], no clear tuning is evident because a heightened response in the 4–6-Hz region was present without stimulation. The subject was quite nervous about the procedure, suggesting that an anxiety tremor may have caused this response. The singers, on the other hand, showed no remarkable peaks without stimulation. This result, although poor in sample size, begins to confirm the

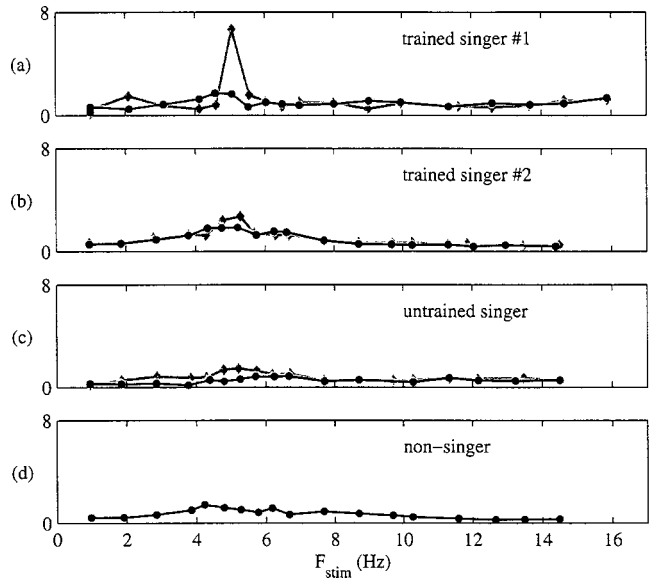


FIG. 8. Resonance curves (amplitude of F_o modulation in Hz) for CT stimulation, with and without natural vibrato. Lines with diamond-shaped data points are for the vibrato case, while lines with the solid black circles are for the *no* vibrato case.

hypothesis that singing training may heighten the reflex gains and establish a reflex resonance. It is important to note that this happened when the singers were deliberately singing a straight tone (vibrato-free).

When the singers allowed their vibrato to be present, the resonance was sharpened (Fig. 8), especially for the trained singer in part (a). The peak rose to 7 Hz (14 Hz peak to peak) and the bandwidth decreased, although exact measurement of the bandwidth was difficult without a finer grid of stimulation frequencies. Such a grid would require a new protocol with new subjects, due to fatiguing factors.

A final data set for the four female subjects (Fig. 9) shows the response curves at different singing pitches: low, mid, and high. The corresponding F_o values were 196 Hz (G3), 330 Hz (E4), and 440 Hz (A4). Only the second trained singer [part (b)] showed a remarkable difference in response to stimulation at different pitches. Lower pitches gave greater response. This subject was a high soprano who probably only engaged her TA muscle at low pitches, thereby producing a better agonist–antagonist reflex loop. The first trained singer [Fig. 9(a)] was a mezzo-soprano who tended to maintain the TA muscle engaged throughout the pitches, which may have resulted in the more uniform patterns. In hindsight, it may have been better to have the subjects perform at *physiologically* equivalent pitches, based on their overall voice-range profile, but given the invasive nature of the experiment it was impossible to repeat the protocol. Also, the voice-range profile would have added to the vocal load, which was already quite heavy.

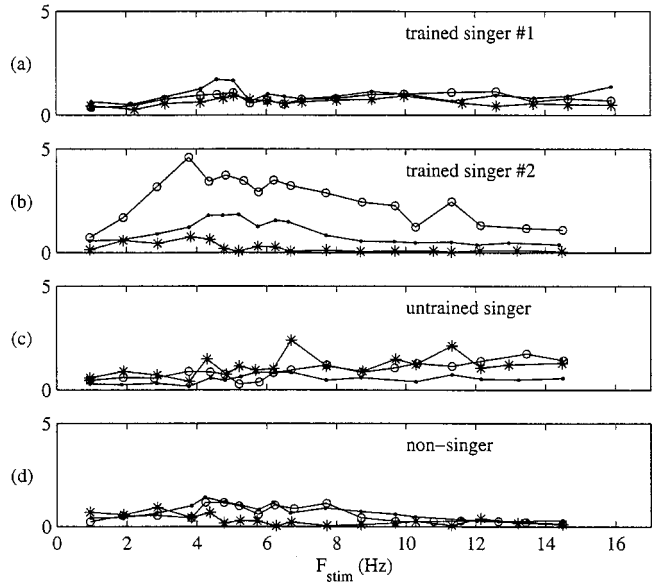


FIG. 9. Resonance curves (amplitude of F_o modulation in Hz) for CT stimulation at three frequencies, no natural vibrato. Open circles represent $F_o = 196$ Hz, solid black circles represent $F_o = 330$ Hz, and asterisks represent $F_o = 440$ Hz.

C. Computer simulation of vibrato

Equations (1)–(15) were programed in MATLAB, using a fourth-order Runge–Kutta technique to solve the differential equations. The computation began by coding the intended muscle memory with a constant level of activation, 20% for CT and 20% for TA, superimposed by either a pulse stimulus or by low-pass-filtered noise in the 0–15-Hz range. That produced the subcortical activities A_{CT} and A_{TA} in Fig. 1. The response to the pulse stimulus is shown in Fig. 10. This figure is essentially a quantification of the hypothetical set of events outlined in Fig. 2. Note the twitch force in the CT following a single pulse activation (top two traces). The pulse was initiated at 0.5 s to allow the tetanic muscle contractions to bring the system to equilibrium in the first half-second. In the third trace from the top, we see the vocal-fold strain exhibiting damped oscillation, a typical response of a natural resonator. The frequency is about 6.5 Hz and the strain variation is less than ± 0.05. The first rise after 0.5 s corresponds to the strain rise hypothesized in Fig. 1(c). Note also that the reflexive TA activity (fourth trace from top) is delayed about 80 ms in reference to the pulse, which was hypothesized in Fig. 1(e). Figure 10 is completed by showing the TA force and the F_o contour in the bottom two traces. Note the further delay between the peaks of TA activity and TA force in the initial excursions.

The F_o contour was computed with the simple vibrating string formula

$$F_o = \frac{1}{2L_o}\sqrt{\frac{F_{\mathrm{TA}}}{\rho S}}, \tag{16}$$

where ρ is the tissue density ($1140\ \mathrm{kg/m^3}$), and L_o and S are the TA muscle length and cross-sectional area as previously defined (Table II). With this formula, the F_o modulation amplitude in Fig. 10 is about 10 Hz initially, damping out with successive cycles. The reflex gains g_1 and g_2 were chosen to be 1.5 in Eqs. (14) and (15), and the neural reflex delays τ_1 and τ_2 were set at 45 ms.

Figure 11 shows the effect of varying reflex gain. An FFT analysis was done on the F_o impulse responses (bottom trace of Fig. 10) for $g_1 = g_2 =$ 1.0, 1.5, and 2.0. The spectra

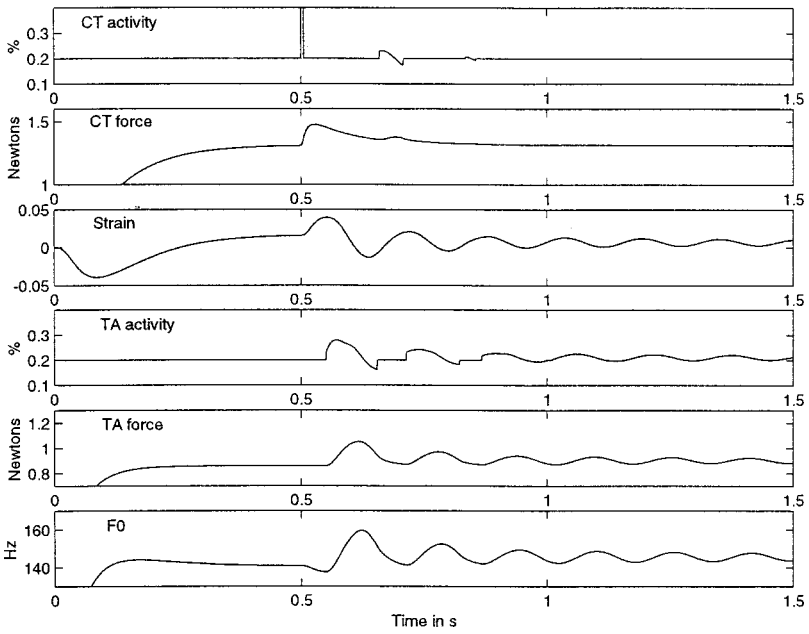


FIG. 10. Simulation of an impulse response of the computer model of vocal vibrato. The reflex gain was set to 1.5.

were normalized to the peak values to show the slight shift in the vibrato frequency with gain. The frequency increases by about 0.5 Hz (from 6.0 to 6.5 Hz) as the gain increases from 1.0 to 2.0. But, much more dramatically, the bandwidth is lowered by increasing the gain. Thus, the tuning to a single vibrato frequency is very sharp for high gain. To put the magnitude of the gain in perspective, a gain of 1.0 means that, according to Eqs. (14) and (15), the feedback (modulation) activity is on the order of 0.02 (the value of the peak strain) superimposed on a background tetanic muscle activity of 0.2. This is a 10% feedback ratio, which the spindles and the sensory-motor synapses have to produce to achieve the oscillation.

It should be pointed out that the normalization of the spectra in Fig. 11 does not give the true picture of increased gain. If normalization were to the noise floor, the gain$=2.0$ curve would rise 30 dB above the gain$=1.0$ curve, a dramatic increase in F_o variation. A lesser, but nevertheless dramatic, increase was seen in Fig. 8(a) for one trained singer when she switched her vibrato on. We reason, therefore, that the reflex gain does not have to change much (only a few percent) to get a substantial increase in vibrato extent.

The second set of simulations was done with low-pass-filtered noise (15 Hz cutoff) to simulate physiologic tremor activity. The result is shown in Fig. 12. The top trace shows constant CT activity ($a_{ct}=0.2$) for the first 0.5 s, followed by low-frequency tremorous activity for the next second. Self-sustained oscillation is seen in the third trace, labeled strain, but the oscillation is not quite sinusoidal. The 6-Hz dominant frequency is clearly visible, however. The gain was set at 1.5, the same as for the pulse excitation in Fig. 10. Note that F_o modulations in Fig. 12 (bottom trace) have an amplitude of about ± 10 Hz, which is $\pm 7\%$ of the mean F_o of 150 Hz. Although this is a little larger than most natural vibrato extents, it is not unrealistic in relation to what trained singer 1 produced in Fig. 8. (Parenthetically, we note that white noise, which was also tried as an excitation, produced smaller F_o amplitudes than the colored low-pass noise.) Figure 13 shows the spectrum of the F_o contour with 0–15-Hz tremor-like noise excitation. The 6-Hz reflex tuning is clearly visible in the low-frequency noise.

VII. DISCUSSION

We now return to the ten observations about vocal vibrato that were listed in the Introduction. The first observation, that vibrato extent is more controllable by a singer than vibrato frequency, is confirmed by our results. When the singers changed voluntarily from vibrato to no-vibrato conditions, the magnitude of the F_o modulation decreased, but the frequency did not change appreciably. Since vibrato extent is controlled by the reflex gains in our model, we suggest that singers adjust reflex gains rather than reflex delays when they change from straight tone to vibrato. The reflex delays are less controllable on a moment-to-moment basis, because they involve neural conduction times, muscle acti-

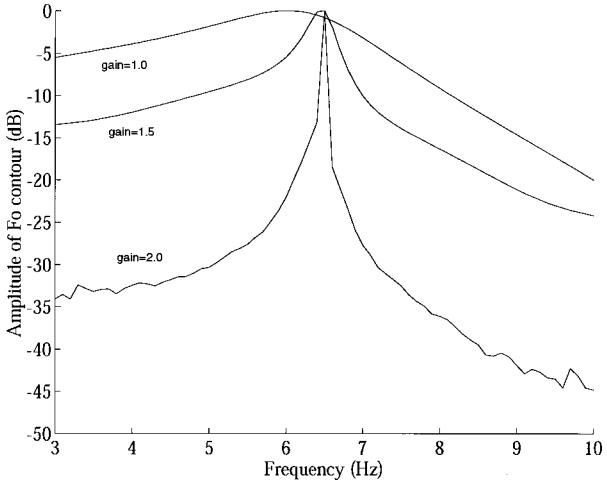


FIG. 11. Frequency spectra of the F_o contours after pulse excitation, for three values of reflex gain. The sampling frequency was 1 kHz and the analysis window was 10 s. The spectra are normalized to their peaks.

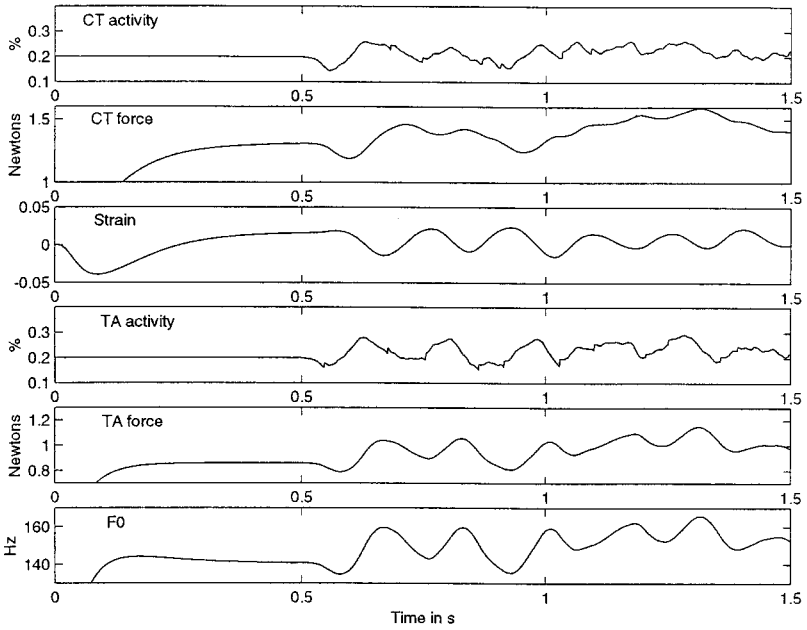


FIG. 12. Simulation of vibrato with tremor-like (0–15-Hz) noise excitation. The reflex gain was 1.5.

vation times, and muscle twitch contraction time, all of which are firm-wired in the neuromuscular system.

The second observation, that many children (and some adults) sing vibrato-free, is also likely to be explained by reflex gains. In children, the thyroarytenoid muscle develops late (Hirano and Sato, 1993), which means that the spindles may not be in place to activate a vocal-fold stretch reflex. In some adults, there is not much length change at all when F_o is varied (Nishizawa *et al.*, 1988). In other words, the CT joint is stiff, or the spaces between the cartilages are small. This may reduce the effective feedback.

The third observation, that some geriatric or poorly conditioned vocalists have a slow and wide vibrato (a wobble), may be explainable on the basis of increased reflex delays. It is well known that nerve conduction velocities slow with age (Chase *et al.*, 1992; Rosenberg *et al.*, 1989; Xi *et al.*, 1999) and that muscle activation and contraction times increase with age. Poor muscle conditioning, regardless of age, also increases the contraction times (slower muscle response). Thus, it is not surprising to see a 1–2-Hz reduction in vibrato frequency with age. A wobble can often be corrected by toning the laryngeal muscles (to increase contraction times). Also, by singing softly at the extremes of the pitch range, less bulk of the TA muscle is involved in vibration, which is likely to reduce the response time.

The fourth observation, that muscularly hyperactive (tense) vocalists sometimes have an excessively fast vibrato (a bleat), may also be explainable on the basis of reflex latencies. Muscles that are already primed (potentiated) respond quicker in a twitch mode than those that are somewhat relaxed. Thus, vocal relaxation exercises (with the mental image of letting the tone flow easily rather than pushing it) could understandably reduce the vibrato frequency. But, the process of change is slow, with significant results often being achieved only over several months of exercising.

The fifth observation, that training reduces the variability (increases regularity) in vibrato rate and extent, may be explained by increased reflex gains, i.e., by a sharper tuning curve. The tremor frequencies are then filtered more effectively to define a narrow band around 5–6 Hz, as was shown with our model.

The sixth observation, that violin, brass, and woodwind vibrato are similar to vocal vibrato, is explainable if we assume that there are always two agonist–antagonist muscles that are involved in similar reflex action. These muscles could be in the hand and fingers (as in violin playing), in the articulatory system, or even in the respiratory system. A respiratory vibrato, however, tends to be slower (around 4 Hz or less) because of the slower response of the muscles around the abdomen, diaphragm, and ribcage. Some mechanical loading may also affect the vibrato frequency. In our model, the mechanical natural frequency for cricothyroid rotation was 17 Hz (based on values of the moment of inertia and rotational stiffness), and the natural frequency of cricothyroid translation was 50 Hz (based on cricoid mass and joint stiffness). These frequencies are too high to mass load the

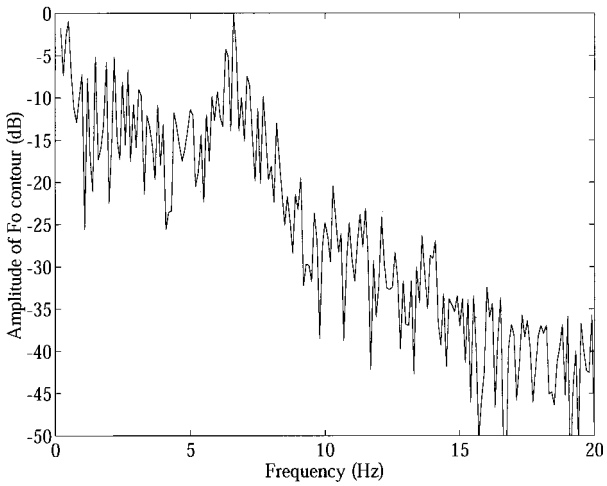


FIG. 13. Frequency spectrum of F_o contour with tremor-like (0–15-Hz) noise excitation. The reflex gain was 1.5.

reflex resonance. But, for larger structures, mechanical loading may be important.

The seventh observation, that some pathological vocal tremors are also around 5–6 Hz (Koda and Ludlow, 1992), suggests that perhaps reflex gains can become too large and no longer able to be inhibited voluntarily. The reflex oscillator is highly underdamped, growing in oscillation amplitude to reach a limit cycle largely determined by structural constraints. This would also explain why boundary structures (jaw, tongue, pharynx) are sometimes engaged in the overall tremulous activity.

The eighth observation, that vibrato frequency increases at the end of a note (Prame, 1994, 1997), may mean that twitch contraction times vary with the degree of muscle length and activation. Perlman and Alipour (1988) showed that twitch contraction times increased with vocal-fold length. As the vocal-folds abduct, the TA muscle is elongated. This may increase the contraction to raise the vibrato frequency, but whether or not it can account for a difference of about 1.0 Hz is not clear. Other explanations should be sought as well.

The ninth observation, that natural vibrato is more easily achieved in the midrange of F_o than at the extremes, may simply mean that the two agonist–antagonist muscles work with nearly equal strains, thereby providing alternating and nearly similar reflex activations, rather than one muscle being highly strained and the other being lax (which provides unequal spindle activation).

The tenth observation, that the natural singer's vibrato can be entrained by an outside stimulus (Titze *et al.*, 1994), has not been a target of investigation here. The auditory reflex has also not yet been included in the current model; it is conceivable that an auditory stimulus from another singer could override the central tremor oscillators and pull the natural vibrato slightly away from its natural resonance, but this awaits further study.

ACKNOWLEDGMENTS

This study was supported by Grant Number R01 DC 04347-01. We are highly appreciative of Jeannie Lovetry, who assisted in this study with funds obtained through a Van Lawrence fellowship from the Voice Foundation. We also thank Professor Charles Larson of Northwestern University, who was instrumental in part of the data collection and design of the mechanical driver for the larynx.

Alipour-Haghighi, F., Titze, I.R., and Durham, P. (**1987**). ''Twitch response in the canine vocalis muscle,'' J. Speech Hear. Res. **30**, 290–294.

Alipour-Haghighi, F., Titze, I.R., and Perlman, A. (**1989**). ''Tetanic contraction in vocal-fold muscle,'' J. Speech Hear. Res. **32**, 226–231.

Alipour-Haghighi, F., Perlman, A., and Titze, I.R. (**1991**). ''Tetanic response of the cricothyroid muscle,'' Ann. Otol. Rhinol. Laryngol. **100**(8), 626–631.

Alipour, F., and Titze, I.R. (**1999**). ''Active and passive characteristics of the canine cricothyroid muscles,'' J. Voice **13**(1), 1–10.

Chase, M.H., Engelhardt, J.K., Adinolfi, A.M., and Chirwa, S.S. (**1992**). ''Age-dependent changes in cat masseter nerve, an electrophysiological and morphological study,'' Brain Res. **586**(2), 279–288.

Cooper, D., Partridge, L., and Alipour-Haghighi, F. (**1993**). ''Muscle energetics, vocal efficiency, and laryngeal biomechanics,'' in *Vocal-Fold Physiology*, edited by I. R. Titze (Singular, San Diego), pp. 37–92.

Damsté, H., Reinders, A., and Tempelaars, S. (**1982**). ''Why should voices quiver?,'' in *vox humana*, studies presented to Aatto Sonninnen. Institute of Finnish Language and Communication, University Jyvaskyla.

Dejonckre, R., Hirano, M., and Sundberg, J. (**1995**). *Vibrato* (Singular, London).

Elble, R., and Koller, W.C. (**1990**). *Tremor* (The Johns Hopkins University Press, Baltimore).

Finnegan, E.T., Luschei, E.S., and Hoffman, H.T. (**1995**). ''Analysis of EMG of intrinsic and extrinsic laryngeal muscles in persons with vocal tremor,'' Paper presented at the annual Voice Foundation meeting, Philadelphia, PA.

Hirano, M., and Sato, K. (**1993**). *Histological Color Atlas of the Human Larynx* (Singular, San Diego).

Hirano, M., Hibi, S., and Hagino, S. (**1995**). ''Physiological aspects of vibrato,'' in *Vibrato* (Singular, London).

Hsiao, T.Y., Solomon, N.P., Luschei, E.S., and Titze, I.R. (**1988**). ''Modulation of fundamental frequency by laryngeal muscles during vibrato,'' J. Voice **1**, 314–319.

Inbar, G., and Eden, G. (**1983**). ''Physiological evidence for central modulation of voice tremor,'' Biol. Cybern. **47**, 1–12.

Keidar, A., Titze, I.R., and Timberlake, C. (**1984**). ''Vibrato characteristics of tenors singing high C's,'' in *Transcripts of the Thirteenth Annual Symposium, Care of the Professional Voice* (The Voice Foundation, New York).

Koda, J., and Ludlow, C. (**1992**). ''An evaluation of laryngeal muscle activation in patients with voice tremor,'' Otolaryngol.-Head Neck Surg. **107**, 684–696.

McMahon, T.A. (**1984**). *Muscles, Reflexes, and Locomotion* (Princeton University Press, Princeton).

Nishizawa, N., Sawashima, M., and Yonemoto, K. (**1988**). ''Vocal-fold length in vocal pitch change,'' in *Vocal-Fold Physiology*, edited by O. Fujimura (Raven), pp. 75–82.

Perlman, A.B., and Alipour-Haghighi, F. (**1988**). ''Comparative study of the physiological properties of the vocalis and cricothyroid muscles,'' Acta Otolaryngol. (Stockh) **105**, 372–378.

Prame, E. (**1994**). ''Measurements of the vibrato rate of ten singers,'' J. Acoust. Soc. Am. **96**, 1979–1984.

Prame, E. (**1997**). ''Vibrato extent and intonation in professional Western lyric singing,'' J. Acoust. Soc. Am. **102**, 616–621.

Ramig, L., and Shipp, T. (**1987**). ''Comparative measures of vocal tremor and vocal vibrato,'' J. Voice **1**(2), 162–167.

Rosenberg, S.I., Malmgren, L.T., and Woo, P. (**1989**). ''Age-related changes in the internal branch of the rat superior laryngeal nerve,'' Arch. Otolaryngol. Head Neck Surg. **115**(1), 78–86.

Sanders, I., Han, Y., Wang, J., and Biller, S. (**1998**). ''Muscle spindles are concentrated in the superior vocalis subcompartment of the human thyroarytenoid muscle,'' J. Voice **12**(1), 7–16.

Seashore, C.E. (**1938**). *Psychology of Music* (McGraw-Hill, New York).

Shiao, T.Y., Solomon, N.P., Luschei, E.S., and Titze, I.R. (**1994**). ''Modulation of fundamental frequency by laryngeal muscles during vibrato,'' J. Voice **8**(3), 224–229.

Stein, R.B. (**1982**). ''What muscle variable(s) does the nervous system control in limb movements?,'' Behav. Brain Sci. **5**, 535–577.

Stein, R.B., and Oguztöreli, M.N. (**1975**). ''Does the velocity sensitivity of muscle spindles stabilize the stretch reflex?,'' Biol. Cybern. **2**, 219–228.

Sundberg, J., Niska-Thörnvik, M., and Söderström, A.M. (**1998**). ''Age and voice quality in professional singers,'' Logopedics, Phoniatrics, Vocology **23**, 169–176.

Titze, I.R. (**1996**). ''Coupling of neural and mechanical oscillations in control of pitch, vibrato, and tremor,'' in *Vocal Fold Physiology, Controlling Complexity*, edited by P. Davis and N. Fletcher (Singular, San Diego), pp. 47–62.

Titze, I.R., and Alipour-Haghighi, F. (unpublished).

Titze, I.R., Solomon, N.P., Luschei, E.S., and Hirano, M. (**1994**). ''Interference between normal vibrato and artificial stimulation of laryngeal muscles at near vibrato rates,'' J. Voice **8**(3), 215–223.

Vilkman, E.A., Pitkänen, R., and Suominen, H. (**1987**). ''Observations on the structure and the biomechanics of cricothyroid articulation,'' Acta Otolaryngol. (Stockh) **103**, 117–126.

Xi, M.C., Liu, R.H., Engelhardt, J.K., Morales, F.R., and Chase, M.H. (**1999**). ''Changes in the axonal conduction velocity of pyramidal tract neurons in the aged cat,'' Neuroscience **92**(1), 219–225.

Coupled modes of the resonance box of the guitar

M. J. Elejabarrieta[a)]
Departamento de Física Aplicada II, Universidad del País Vasco, Apdo. 644-48080 Bilbao, Spain

A. Ezcurra[b)]
Departamento de Física, Universidad Pública de Navarra, 31006 Pamplona, Spain

C. Santamaría
Departamento de Física Aplicada II, Universidad del País Vasco, Apdo. 644-48080 Bilbao, Spain

(Received 1 November 2001; revised 5 February 2002; accepted 16 February 2002)

Vibrations of the resonance box of the guitar have been studied by means of the modal analysis technique and the finite-element method. An expert craftsman constructed the guitar box with all the structures, internal and external, characteristic of a real instrument for the experimental measurements. The boundary conditions were chosen in order to clarify the soundboard–back interaction only via the internal air coupling. The numerical model allows one to study the influence of each component on the whole box, and the contribution of the modes of the components (wooden box and its parts, and air), to the coupled modes by calculating their participation factors. The coupled modes of the guitar box are discussed taking into account both the finite-element and modal analysis results. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1470163]

PACS numbers: 43.75.Gh [ADP]

I. INTRODUCTION

The guitar is a complex mechanical system because its dynamic behavior is determined by the interaction of several components. The plucked strings radiate only a small amount of sound directly, but they excite the bridge and the top plate, which in turn transfer energy to the air cavity, ribs, and back plate. Sound is radiated efficiently by the vibrating plates and through the sound hole. In particular, two main distinguishable parts form the guitar box: the wood structure and the inside fluid. This paper is devoted to the experimental and numerical determination of the vibrational behavior of the guitar box and forms part of a more extensive study addressing the vibrational behavior of the guitar and its pieces.

Several experimental and computational techniques have been applied to this issue. Holography and laser interferometry have provided very accurate information about guitar-box resonances[1–4] both in classic and acoustic guitars, but are devoted mainly to the soundboard alone. Regarding the complete resonance box, Richardson[5] analyzed its modal parameters under free-boundary conditions by means of holographic measurements. Another two techniques, the finite-element method and the modal analysis technique, are quantitative and complementary, since their results can be compared. The modal analysis technique allows an approach to the sound produced by the instrument, since the vibrational analysis is carried out on the complete body of the guitar: structure and fluid. By contrast, this type of analysis is not able to study the influence of each part of the guitar on the complete instrument in a direct way, and must be applied together with other techniques; for example, numerical calculation by the finite-element method. This latter method has become an important tool for studying the vibrational behavior of complex mechanical systems, such as musical instruments.[5–8] In this sense, the dynamics of the guitar soundboard both in free- and fixed-boundary conditions has been studied. The modal analysis technique was applied by Marshall[9] to determine the vibrational behavior of the violin. In the case of the guitar box, some studies have been published previously on electromechanical input–output measurements at several points of the soundboard and their correlation with acoustic efficiency,[3,10,11] although the modal parameters of the system were not determined.

Here, we report the calculated low-frequency modes and natural frequencies of the resonance box (top plate, back side, ribs, edges, and blocks) starting on the dynamics of the fundamental components: the top and back plates. The neck was not included. The boundary conditions were chosen in order to make the soundboard–back interaction clear exclusively through the internal air. First, the behavior of the complete box was obtained in the absence of air starting from the soundboard and back plate. This numerical model allowed us to evaluate the influence of the constructing process of the top and back plates on the definitive instrument. Then, the air inside the cavity was taken into account and the coupled modes were calculated. This computer calculation permits, on the one hand, a quantitative study of the interaction between the fluid and the structure, and on the other, to assess the influence of the constructing process of the top and back plates on the dynamics of the instrument.

The numerical model corresponds in detail to a real guitar box, as designed and constructed by a skilled craftsman. The modal analysis technique has been applied to determine the natural frequencies, vibration modes, and quality factors of this resonance box, under equivalent boundary conditions. Finally, the numerical and experimental results have been compared and discussed.

[a)]Present address: Departamento de Mecánica, Mondragon Unibertsitatea, 20500 Mondragon, Guipuzcoa, Spain.
[b)]Electronic mail: aezcurra@unavarra.es

II. EXPERIMENTAL SETUP

A. The guitar box

The system under study was a guitar resonance box. On the one hand, it was constructed as a ''real'' system, and on the other, as a numerical model. The materials and the construction process were decided by a skilled luthier. The top plate was made of Canadian cedar and the thickness varied from point to point (2–2.7 mm). Some additional structures were provided to the plate. In particular, two bars were placed perpendicular to the grain, respectively, over and below the sound hole. The upper bar had a constant rectangular cross section. The lower bar also had a rectangular section, but its height was reduced linearly from the two ends to the center of the bar. Seven fan struts were arranged in the classic way, according to the *Torres* distribution and two symmetrical struts were placed on the lowest part of the plate. The bars and fan struts were made of spruce. The final mass of the soundboard with the whole set of internal structures was 165.2 g. The dynamics of the soundboard was previously analyzed in the successive stages of its construction.[12]

The back plate was made of Indian rosewood (thickness 2 mm) with one longitudinal and three transversal spruce bars added. The final mass was 281.0 g. The ribs were also made of Indian rosewood (thickness 1.7 mm) and their height increased linearly from 97 mm in the upper zone to 105 mm in the lower part. The top and back plates were joined to the ribs along their contour by means of the edges or *junquillos*, rectangular pieces of 3×10 mm. Two blocks were added, in the lower and upper zones, to avoid the strain of the box. The tail block had a uniform, nearly rectangular cross section of 10×50 mm. The upper block had a width of 60 mm, and its length increased linearly from 75 mm at the top plate to 90 mm at the back. Both edges and blocks were made of spruce. The final mass of the box was 853.0 g.

Among other possibilities, the finite-element method allows one to obtain the vibrational behavior corresponding to complex mechanical systems through a computational simulation that simplifies the continuous system into a discrete system. The data required are the geometric size and shape, the boundary conditions of the piece, and the characteristics of the materials: density, elastic tensor, and damping. Here, we attempted to achieve a numerical model of the guitar box with a high degree of accuracy as regards both the geometrical design and the material parameters. A partial model was previously used to study the soundboard dynamics along its construction process.[13] It should thus allow us to study the effects of several parameters on the behavior of the soundboard, both alone and when forming part of the box.

The numerical model evolved following the real manufacturing process designed by the craftsman, by modifying its geometric characteristics, and by adding the internal structures, and became quantitatively more complex as the procedure progressed. Figure 1 shows two images of half of the mesh defined for the resonance box of the guitar. The top plate was modeled with second-order brick elements (6590 nodes and 956 elements). The back plate was modeled with second-order brick elements (7362 nodes and 1116 elements). The ribs were modeled with second-order sheet elements (1938 nodes and 612 elements). The bridge was modeled with first-order brick elements (116 nodes and 40 elements). The rest of the pieces, blocks and edges, were modeled with first-order brick elements, as were the bars and rods of the soundboard and back plate. The complete resonance box contained 15 946 nodes and 3132 elements. The resolution of the mesh guarantees the response up to 600 Hz.[14] The material parameters used for the numerical model are shown in Table I. No viscous effects were considered, and hence the modes for both the independent domains and the coupled system were normal.

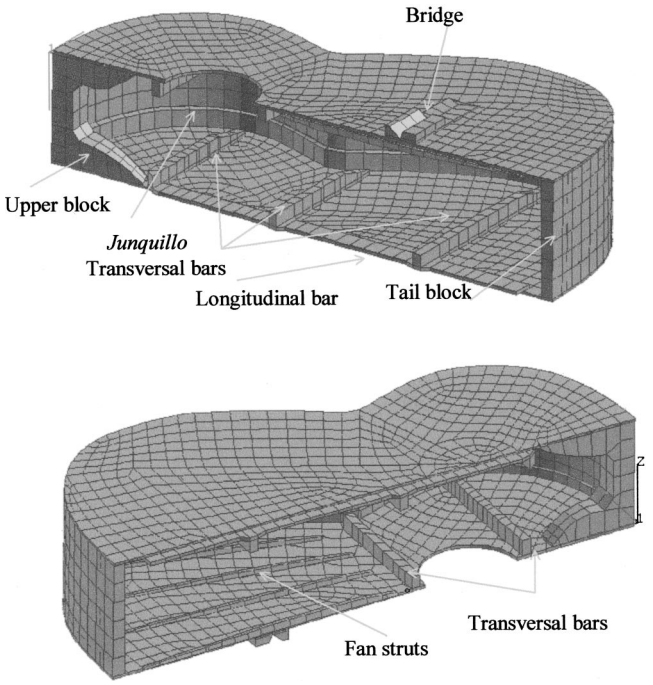


FIG. 1. Two longitudinal sections of the finite element mesh; the geometrical design and the added structures can be seen.

The mesh and the boundary conditions for the air inside the cavity have been described together with its dynamic behavior when the cavity is completely rigid in Ref. 15. The mesh had 8474 first-order cubic elements from the ABAQUS[14] library, specific for acoustic media, and 10 788 nodes. The geometrical outline of the surface of the cavity included the internal struts. The values for the bulk modulus B and the density ρ of air are as follows: $B=142$ kPa and $\rho=1.2$ kgm^{-3}, whereas viscous effects were not taken into account.

B. Numerical calculations

Regarding the numerical model, the analysis, and the interpretation of the results, the fluid–structure coupling is a

TABLE I. Elastic parameters and densities corresponding to the components of the guitar box as explained in the text.

Elastic constants	Canadian cedar	Indian rosewood	Spruce
E_1 (MPa)	5600	16 000	15 000
E_2 (MPa)	520	2 200	650
G_{12} (MPa)	600	1 100	600
ν_{12}	0.33	0.36	0.03
Density ρ (kg m^{-3})	342	775	460

complex phenomenon. The coupling takes place at the interface of the two domains through the imposed boundary conditions. These domains describe different physical situations, but none of them can be analyzed without taking into account the influence of the others. In our case, the interaction was limited to small amplitude movements. The application of the finite-element analysis was accomplished using the ABAQUS software[14] implemented on an Alpha 2100 workstation, in the case of the wooden box and the inner air individually. The coupled numerical model and application of the finite-element analysis were accomplished using the SYSNOISE software,[16] since the ABAQUS software does not incorporate the algorithm of coupled modes analysis. On the other hand, SYSNOISE software only supports plate elements with homogeneous and isotropic materials, while ABAQUS has a wider element bookstore and options to describe different material behaviors, in particular orthotropic materials. Because coupling analysis and anisotropic materials were involved, both programs were necessary and all the information had to be transferred between them. Moreover, the data had to be adapted not only in their address format but also in their normalization factors. The results were renormalized from the previous unit amplitude normalization to modal mass unit by using the adequate normalization factors. After calculating the vibration modes of both the wooden box and the air, they were combined by means of the modal coupling method,[16] which is based on the overlap and expansion of the vibration modes of the two uncoupled systems. Since the vibration modes of a system contain all the information about it, it is feasible to define such a system by its modal parameters (eigenvectors and eigenvalues) instead of defining it through its mechanical parameters (mass, stiffness, and damping). Regarding the boundary conditions, the ribs were considered as completely clamped to avoid their displacement, reproducing the experimental conditions of the modal analysis described afterwards.

Thus, the dynamic behavior of the resonance box of the guitar in the low-frequency range was calculated through the vibration modes of the structure and of the fluid inside the cavity, with the ribs fixed. On these uncoupled modes, the modal coupling method was applied to calculate the vibration patterns and natural frequencies of the coupled modes of the box. The interface is the air surface in contact with the inner surface of the box. Both surfaces are similar, but the nodes and elements do not coincide completely and hence the displacement of the structural nodes corresponding to the vibration modes of the air was interpolated on the nodes of the interface. Although several approximations are involved in the numerical method, the results are in fair agreement with the experimental measurements and allow us to analyze them.

C. Experimental measurements

The experimental method was the modal analysis technique.[12,17] Among the available applications, the frequency response function (FRF) was chosen. The structure was put into vibration by a transducer hammer acting on successive points, and the response was recorded at one fixed point (the so-called roving hammer method) through an accelerometer. More details can be found in Ref. 12. The frequencies studied were in the 0- to 800-Hz range, whereas the resolution was 1 Hz, although the estimation of the modal parameters was carried out in more reduced bands. Both the excitement and the response were perpendicular to the plates, the most important direction regarding the vibration and acoustic radiation of the final instrument. The response point was located at the position of the bridge, next to the bass string, and 230 excitation points were distributed over the soundboard (115) and the back (115). The number and location of the selected measurement points (maximum distance 5.85 cm; minimum distance 2.82 cm) covered the frequency range as well as the number of expected modes. The estimation of the modal parameters was global, since each modal analysis was made up of the 230 response functions. Therefore, the reported natural frequencies, the vibration modes, and the quality factors do not depend on the particular response point. The vibration modes obtained had a normal character with proportional viscous damping. The most relevant characteristics of the data acquisition and analysis can be found in Ref. 12. As for the boundary conditions, the ribs were fixed by means of polyurethane foam to a metallic mold, in order to prevent the soundboard and back-plate perimeters from moving in the analyzed frequency band.

III. RESULTS AND DISCUSSION

A. The soundboard and the back plate by the finite-element method

The starting point is the behavior of the main components of the box, soundboard and back plate, individually, under hinged boundary conditions; that is, allowing the edge to have zero displacement and free slope, the so-called simply supported boundary conditions. In this stage both components were finished with the complete strut system (bridge, transverse bars, and fan struts in the case of the soundboard, and longitudinal and transverse bars in the case of the back). Figure 2 shows the six lowest modes corresponding to each component, soundboard and back plate. The modes are named as (m,n), m being the number of transverse half-waves and n the number of longitudinal half-waves on the plate,[5] adding an ''*s*'' or a ''*b*'' to refer to the soundboard or to the back, respectively. As can be seen, the patterns are governed by the presence of the main struts. In the case of the soundboard, the position of the bridge determines the antinodal zones below the sound hole. In the case of the back, the transverse bars prevent the transversal flexure patterns from appearing in the low-frequency range and the antinodal zones are of small area and situated between the bar positions. Concerning the natural frequencies, the greater stiffness of the material of the back (rosewood), together with its stronger strut system, makes the frequency band broader in this component than in the soundboard.

B. The empty box by the finite-element method

In this section the eight lowest modes corresponding to the empty resonance box are presented. The soundboard and the back plate have now been attached to the ribs along with *junquillos* and blocks forming the ''empty box;'' that is,

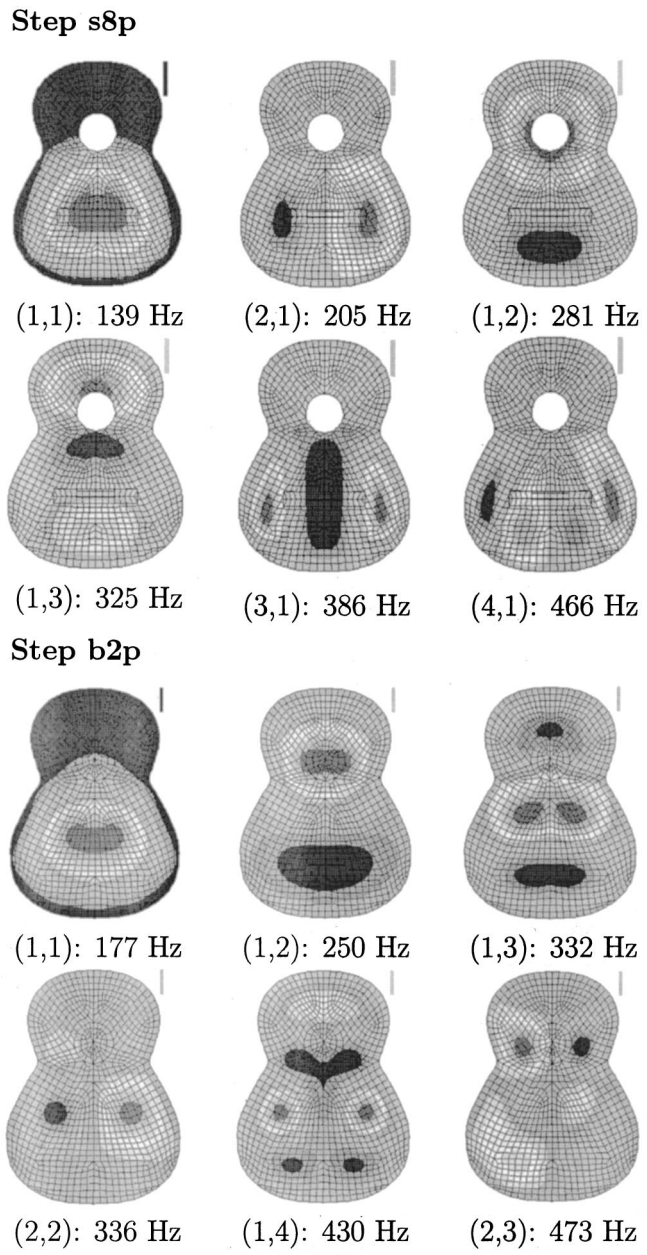


FIG. 2. Calculated modes of vibration and natural frequencies of the soundboard and the back in hinged boundary conditions. The stripe near each mode indicates the nodal zones color.

when the air is away. The patterns were normalized to unit amplitude, and were obtained by fixing the ribs; that is, imposing zero displacement and zero slope conditions. Each vibration mode is represented by two images, one corresponding to the soundboard and the other to the back plate. The two-dimension representation was chosen for clarity. Figure 3 shows the eight lowest modes of the empty resonance box. The frequency band is up to 500 Hz, covering more than half of the fundamental tones of the instrument. As can be seen, the vibration appears only in one part (soundboard or back) but not in both simultaneously. This result is obvious: if the ribs are fixed and the cavity is empty, there is no way to couple these two parts, and the modes are limited to only one of them. The names of the modes are as in the preceding stage, when the components were independent, although the shapes of some of them have changed.

If the soundboard and the back were identical both in design and material, two identical modes would appear at each natural frequency, one in each component. Not even in this case would coupling be present since there is no way of transmission, structural or acoustic, between the components. In our case, to a large extent the eight modes correspond to the vibration modes of the top (4) and back (4) plates with their contours fixed, although they are mainly affected by the blocks, the upper block (due to its size and position) being the most influential one. In the case of the back, the present modes are the fundamental mode $(1,1)b$, the longitudinal flexion modes $(1,2)b$ and $(1,3)b$, and the shear mode $(2,2)b$. In the case of the soundboard, the patterns are the fundamental mode $(1,1)s$, the transverse flexural modes $(2,1)s$ and $(3,1)s$, and the longitudinal mode $(1,2)s$.

Figure 3 suggests that some remarks should be made.

(i) The vibration area is slightly greater in the soundboard than in the back for the (1,1) mode, and is located in the lower zone below the sound hole while the upper zone remains motionless. The amplitudes are not comparable since each mode has been normalized to unit amplitude.

(ii) The aspect of mode (1,2) is similar in both components in the lowest antinodal zone. In the case of the upper maximum, it is situated in the zone of the sound hole for the top plate, and expanded by the two transverse bars in the case of the back.

(iii) Although mode $(3,1)s$ resembles a $(3,2)s$ mode, it comes from the (3,1) mode of the soundboard with the contour fixed. The central antinodal zone is split into two, one of them below the bridge and the other one over the zone of the sound hole.

(iv) Mode $(1,3)b$ presents a similar split; in this case it resembles a $(2,3)b$ mode.

(v) In the case of modes $(2,1)s$ and $(2,2)b$, their aspect is similar to that of the individual components with their contours fixed.

Concerning the natural frequencies, the only significant changes appear in the modes affected by the assembly of the box. In this case the modes are more rigid, since their frequencies increase in comparison with the results of the preceding section. The frequencies most sensitive to the assembly procedure are those of the patterns of the back $(1,2)b$ and $(1,3)b$, with a relative increase of 15% with respect to the preceding stage. In the case of the top plate, the modes sensitive to the assembly procedure, $(1,2)s$ and $(3,1)s$, increase their frequencies by less than 7%.

In view of the above, it is possible to predict the behavior of the resonance box by observing the dynamics of the independent components under fixed boundary conditions. The only modes affected are those presenting vibration in the zone of the upper block. This piece hinders the movement in its zone and increases the frequencies of the affected modes.

C. The air inside the box by the finite-element method

As mentioned above, the vibration modes of the enclosed air have been determined in a previous paper.[15] To

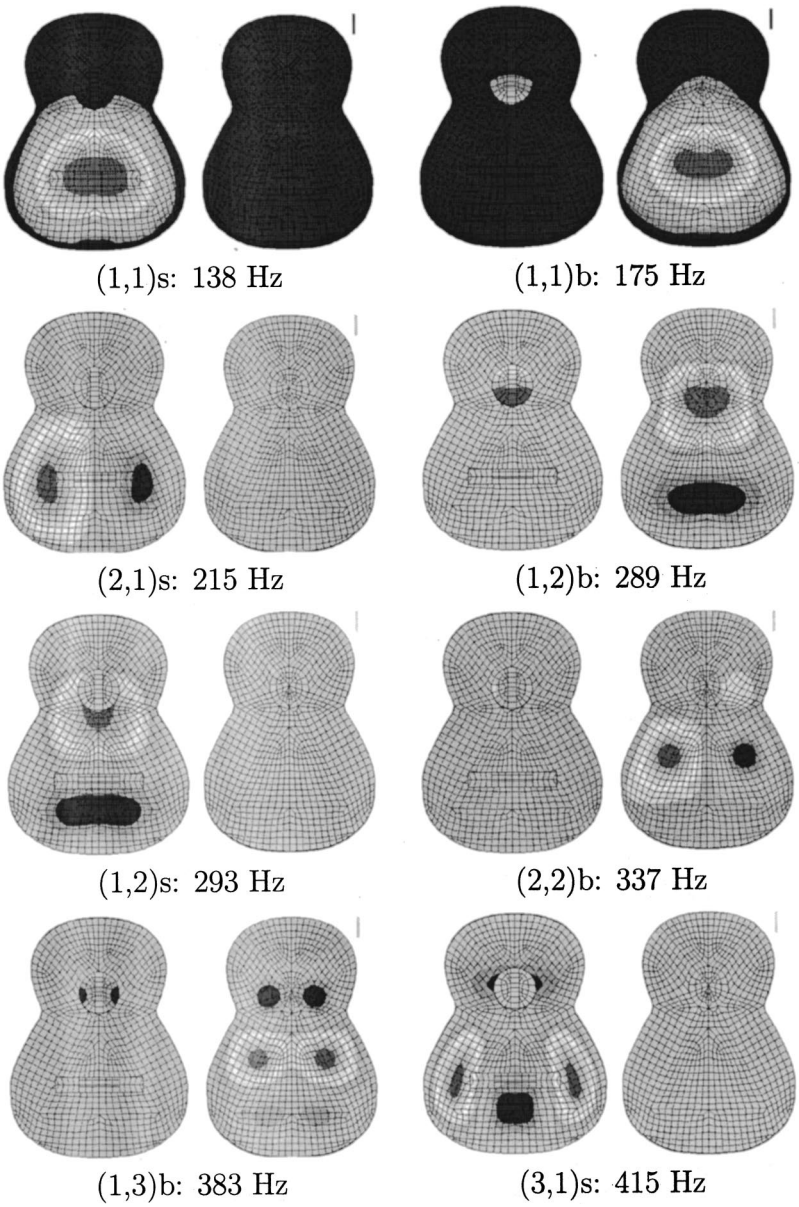


FIG. 3. Calculated modes of vibration and natural frequencies of the empty guitar box with the ribs fixed. The stripe near each mode indicates the nodal zones color.

resume, the fluid geometry corresponding to the guitar box, including internal bars and fan struts, was simulated. A preliminary finite-element model was used in order to define the boundary conditions on the sound hole so that the obtained internal modes were equal to those corresponding to the rigid box immersed in a much bigger space with free-boundary conditions on the external surfaces. The six lowest patterns were similar to those experimentally determined by other authors.[18–20] The calculated natural frequencies are presented in Table II, named in the usual way, starting from A0 (the so-called Helmholtz resonance).

TABLE II. Natural frequencies of the air cavity of the guitar as determined by the finite-element method (Ref. 15).

Modes	Frequencies (Hz)
A0	155
A1	418
A2	545
A3	718
A4	771
A5	981

D. The guitar box: Numerical calculations and experimental measurements

This section addresses the dynamics of the complete guitar box; that is, the coupled modes of the wood structure together with the inside air; these are indexed by the capital letters TB, referring to the top–back coupling, followed by the mode number, starting at the lowest frequency.

Figure 4 shows the eight lowest coupled modes, as determined by the finite-element method; since viscous effects were disregarded both in the wood structure and in the fluid, the coupled modes are normal. The boundary conditions were the same as in the preceding paragraph; that is, fixing the ribs. For graphic representation, the vibration amplitudes in both the soundboard and in the back show the coupled modes; for clarity's sake the additional struts have been

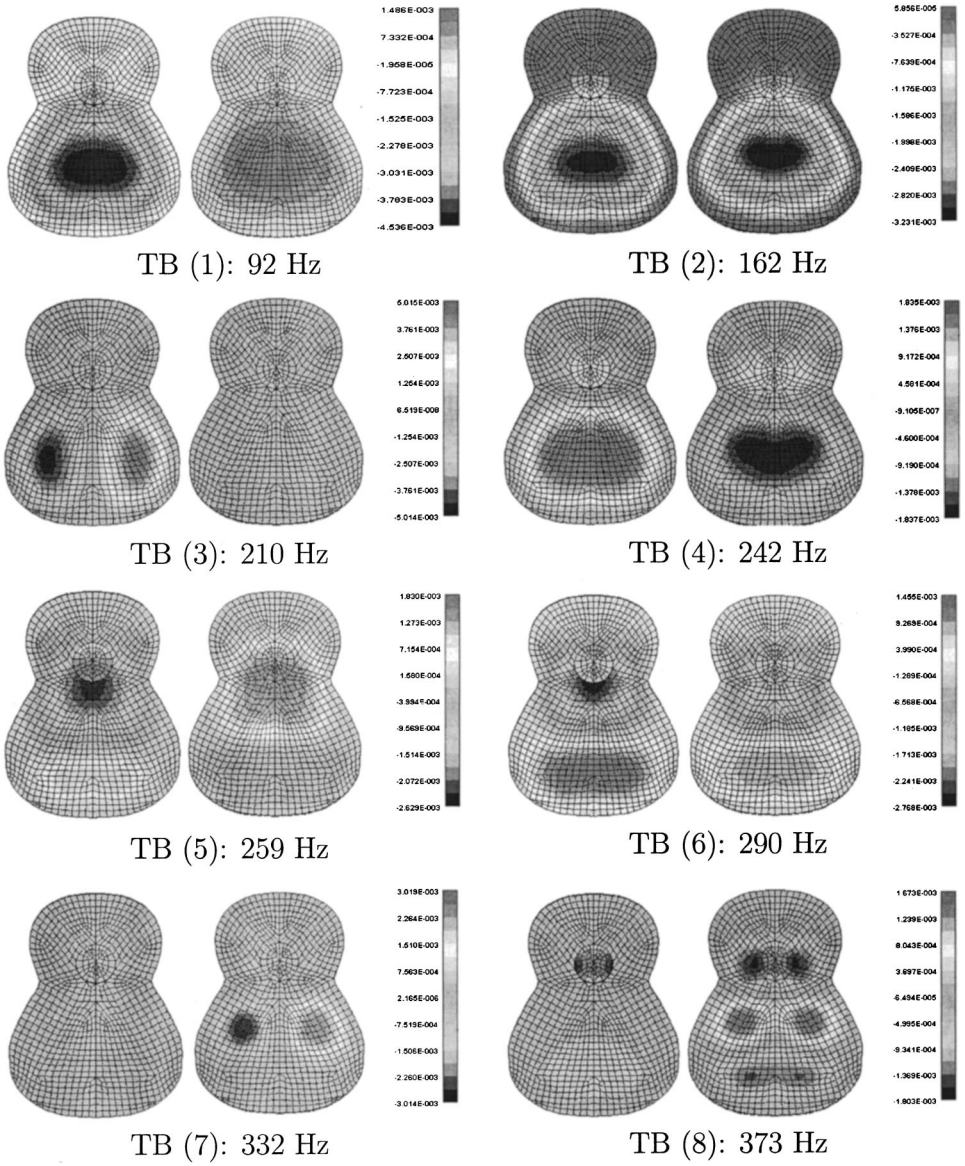


FIG. 4. Calculated coupled vibration modes and natural frequencies in (Hz) of the resonance box of the guitar with the ribs fixed. The back is viewed as if the soundboard were transparent. The colored stripe indicates the amplitude scale. The amplitude values cannot be compared with one another because they are independently normalized.

omitted in the figure, although they were taken into account in the whole analytical and experimental process. As can be seen, the air inside the cavity substantially affects the vibration patterns. Thus, it causes the soundboard and the back to move together at certain frequencies. Up to now both top and back plates had vibrated independently because there was no elastic medium for them to interact. The air contained in the box allows the connection between both structural components. In this way the coupling among the soundboard, the air, and the back gives rise to the modes of the box.

The calculated modes show that the influence of the air is greater at lower frequency. On comparing this figure with Fig. 3, it may be seen that the natural frequencies of the corresponding structural modes have decreased; this means that the fluid acts as an added mass. Moreover, the interaction between the structural and acoustic modes depends on the vibrational character of the structural modes. For example, some modes only appear in one component—top or back—i.e., the case of the transversal flexural modes.

Figure 5 shows the seven lower coupled modes, their eigenfrequencies and quality factors, as determined by the modal analysis measurements. The vibration amplitudes are not comparable, since each pattern has been normalized to unit modal mass. The gray zones indicate immobility. Thus, from the figure it can be seen how the external edges of both the soundboard and the back remained fixed. This guarantees that the boundary conditions imposed on the ribs of the instrument prevent their movement and that the modes both in the soundboard and in the back are originated by the coupling of these structural components via the fluid of the cavity. Each vibration mode was estimated from the 230 responses of both plates. Because of this, each pattern in Fig. 5 represents the vibration in both components. Under these conditions, it is possible to compare the contour line density in the soundboard and in the back for each vibration mode. Thus, the number of contour lines in each component will be indicative of the amplitude of vibration in these structures. As can be seen, most of the largest amplitudes are situated in the lower and middle zones of the box in the analyzed frequency band.

As can be seen by comparing Figs. 4 and 5, the proposed finite-element model predictions are in fair agreement with

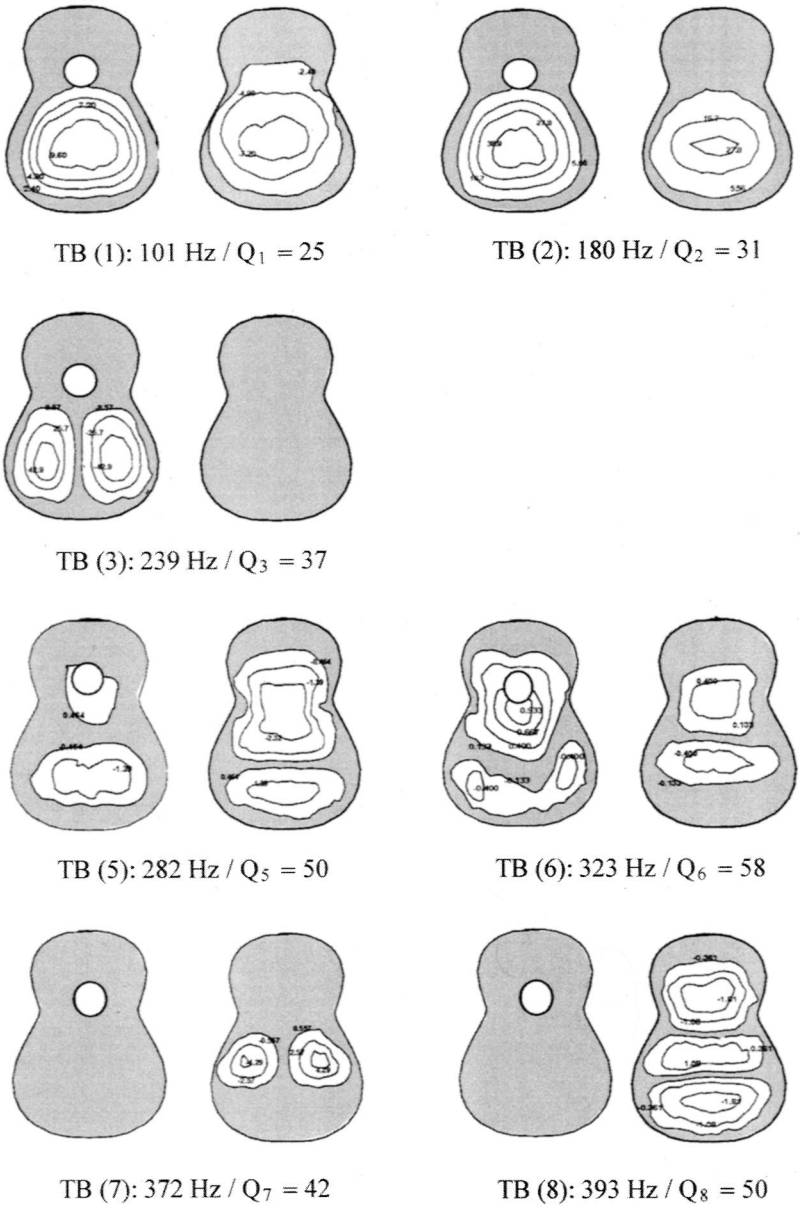


FIG. 5. Modes of vibration and natural frequencies of the resonance box of the guitar with the fixed ribs as obtained from experimental measurements analyzed by the modal analysis technique. The back is viewed as if the soundboard were transparent.

the experimental dynamics of the guitar box in the low-frequency range.

Figure 6 shows the modal participation factors of the uncoupled modes in the coupled modes. A pyramidal representation was adopted for the structural modes, and pie charts for the acoustic patterns. Furthermore, each mode is represented by a different fill pattern to facilitate identification. Contributions of less than 5% were disregarded. This participation index allows one to establish the origin of the coupled modes and to predict the necessary modifications in the soundboard and the back to obtain the desired effects on the definitive resonance box. The following sections are devoted to the analysis of the modes presented, accounting for the modal participation factors.

1. TB(1)

This is the fundamental mode of the resonance box. In this pattern the soundboard presents an antinodal zone at the position of the bridge. The upper zone of the soundboard remains motionless due to the presence of the transverse bars and the upper block. Also in this case, vibration is limited to the lower part of the back but in this case the amplitude of vibration is three times smaller than in the soundboard in the calculated mode; the experimental results only indicate that the vibration amplitude is higher in the soundboard. Figures 4 and 5 show that the top-and back plates vibrate in phase opposition; this leads to considerable changes in the inner volume. Therefore, it is clear that the TB(1) mode is an efficient mode of radiation in the low-frequency band.[21] TB(1) is originated from modes $(1,1)s$ and $(1,1)b$ of the empty resonance box (see Fig. 3), mainly interacting with the A0 air mode, as seen in Fig. 6. Mode A1 is also present, but its influence is negligible in comparison with the A0 mode. The principal structural mode is the soundboard's fundamental $(1,1)s$, with a participation factor of 70%. Thus, any modification to the soundboard, either the wood or the struts, would strongly affect the frequency and shape of this mode. In the case of the fluid, this mode is addressed by the Helm-

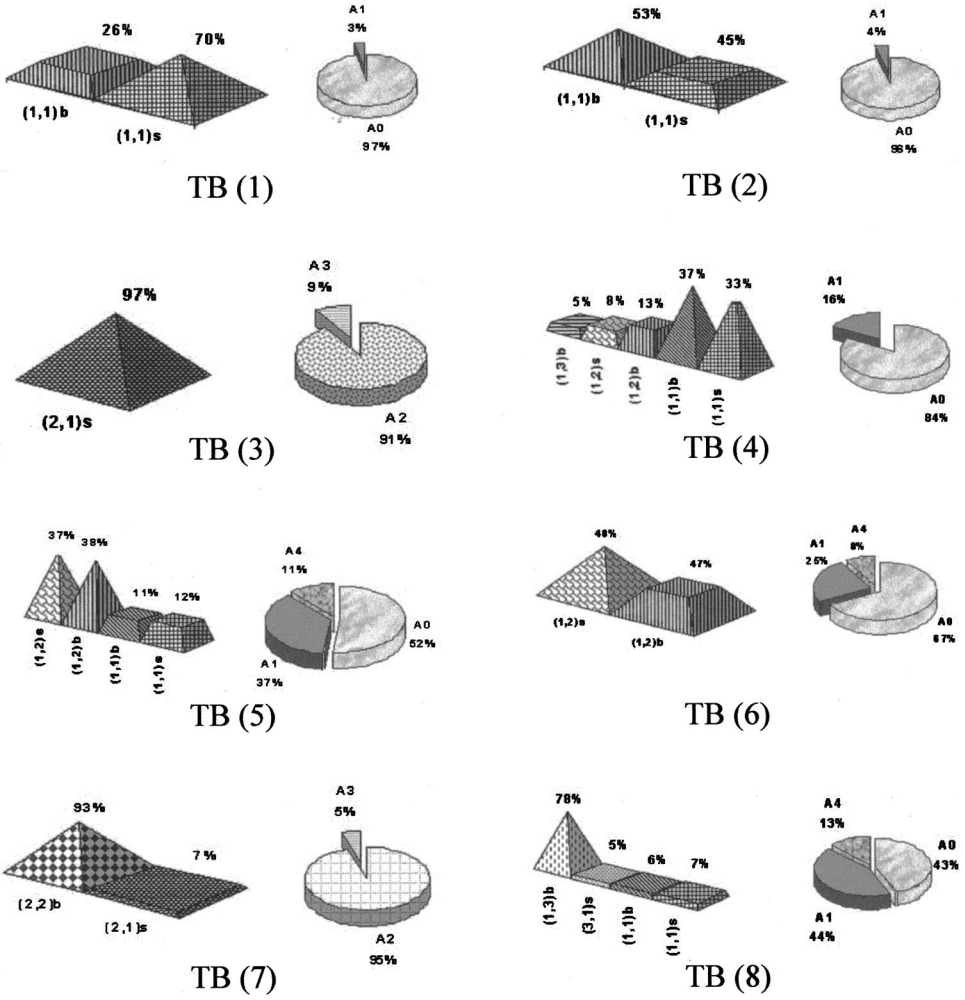


FIG. 6. Participation factors in (%) of the structural and acoustic vibration modes.

holtz resonance, and hence the height of the ribs or the size and shape of the sound hole would affect this fundamental mode of the resonance box. The frequency is reduced by 40% with respect to the empty box, pointing to the added mass effect of the fluid interacting with the structure.

2. TB(2)

The second vibration mode of the resonance box is named T(1,1) or first top plate resonance by other authors. The shape of this mode is similar to the TB(1) mode but in this case soundboard and back vibrate in phase, and so the involved volume changes in the inner air are smaller. Figure 5 shows that the maximum displacement of soundboard and back are equal and situated in the lower part of the box. Just as the TB(1) mode, this pattern is originated by the interaction of the fundamental modes of the soundboard $(1,1)s$ and back $(1,1)b$ with the Helmholtz mode A0. However, in contrast to mode TB(1), in this case the participation of the two modes of the empty box is nearly equal (see Fig. 6), hence, this mode will be equally sensitive to structural modifications in both the soundboard and the back plate. The $(1,1)s-(1,1)b$ coupling is mainly carried through the fundamental resonance of the air cavity A0, although mode A1 is also present. The frequency decrease with respect to the empty box highlights the added mass effect of the fluid, just as in the TB(1) mode. The importance of TB(2) has been made clear by other authors, who have shown that it is the most efficient mode of radiation in the low-frequency band.[3,21]

Both TB(1) and TB(2) have been analyzed by other authors[19,21–24] experimentally and by means of simplified models. The theoretical models are able to justify the existence of these modes, but not to determine the proportions in which each component contributes to the coupling.

3. TB(3)

In this mode the back remains motionless: as has been explained before, the internal bars of the back hinder the vibration of its transversal flexural patterns and thus the contribution to radiation due to the back will be negligible. No changes were found on comparing TB(3) with $(2,1)s$, either in pattern or in frequency, and therefore no soundboard–back coupling is present in this mode, so it can be analyzed as a soundboard mode. It will be sensitive, both in frequency and in the vibration amplitude, to the placement of the transversal bars and fan struts close to antinodal zones.[13] The most influential acoustic mode is A2, for it presents a central longitudinal nodal line just like mode $(2,1)s$.

4. TB(4)

Its pattern is similar to TB(1) mode, involving large volume changes, so TB(4) is an efficient mode of radiation, too.

It is originated from modes $(1,1)s$ and $(1,1)b$ of the empty box, mainly interacting with the A0 air mode. As the participation factors are similar to TB(1), the same considerations on the influence of the soundboard and the geometry of the cavity can be made in this case. Attending to this behavior, we can infer that this mode forms the low-frequency triplet together with TB(1) and TB(2);[25,26] the calculated TB(4) is more complex than the theoretically predicted mode as the structural modes $(1,2)b$, $(1,2)s$, and $(1,3)b$ are present too. This mode could not be experimentally determined, probably due to the proximity of TB(3).

5. TB(5)

The soundboard and the back vibrate in phase opposition, with a longitudinal flexion character. The maximum vibration amplitude occurs around the sound hole in both plates. The lower part of the box, below the bridge, displays a weak movement that is more visible in the soundboard. The experimental measurements indicate that the vibration amplitude is larger for the back than for the soundboard; this fact is not so clear in the calculated mode. On analyzing this pattern (see Fig. 6), some remarks are merited. Regarding the wood structure, the soundboard and back contribute to the same extent, both of them providing two modes. Furthermore, the main modal participation is due to modes $(1,2)s$ and $(1,2)b$ and is three times higher than the participation of the corresponding fundamental modes. This mode is therefore governed by the longitudinal patterns of the soundboard and back.[13] It will be sensitive to any change in the soundboard fan and to the distribution and shape of the transverse bars of the back. Regarding the acoustic modes, three patterns appear: A0, A1, and A4. Although mode A0 is still predominant, as in TB(1) and TB(2), here A1 and A4 have a notable participation in the coupling: also in this case, the structural–acoustic coupling is present when the modes have the same vibrational character. The participation of modes A1 and A4 implies that mode TB(5) is governed by the stationary waves of the air cavity, and hence the internal shape of the resonance box will influence this mode both through the structural response and through the air cavity response. This mode, like TB(1) and TB(2), is a strong radiator[19] and together with the TB(6) mode covers most of the fundamental notes and most of the second partials of the lowest notes of the instrument.

6. TB(6)

This pattern presents a character of longitudinal flexion, just like TB(5), but in this case the soundboard and back vibrate in phase, so the air volume changes are smaller. Figures 4 and 5 show that the maximum amplitude happens at the soundboard, around the sound hole. Unlike mode TB(5), this mode presents a notable movement around the bridge and will therefore be easily excited by the strings. Figure 6 shows the structural and acoustic modes forming this pattern. The two structural modes are the longitudinal patterns $(1,2)s$ and $(1,2)b$ of the empty box (see Fig. 3) in equal proportions. The acoustic modes are A0, A1, and A4, as in TB(5), although the influence of A0 is higher in TB(6). Regarding the frequency (290 Hz), this is almost equal to the frequency of the structural modes of the empty box (289 and 293 Hz).

Modes TB(5) and TB(6) form a symmetrical–antisymmetrical pair, as do TB(1) and TB(2). Both modes vibrate with a longitudinal flexion character with two antinodal zones in the lower and middle zones of each component. The presence of vibration all around the sound hole indicates that both modes will be efficient in sound radiation.

7. TB(7)

In this mode the soundboard remains motionless, just as the back did in TB(4). Both TB(4) and TB(7) have a transversal flexural character. Unlike mode TB(4), in which no back modes were present, in TB(7) the soundboard mode $(2,1)s$ certainly appears (participation 7%; see Fig. 6), although no soundboard vibration is evident in Figs. 4 and 5. The acoustical modes are the A2 mode (95%) and A3 (5%). Just as in mode TB(4), the air modes are those with a (zero-pressure) nodal line along the longitudinal axis of the box. Since this mode appears only in the back without coupling with the soundboard, acting on the strings will not excite it and so it does not contribute to the sound radiation and quality of the guitar.

8. TB(8)

This mode has a longitudinal flexural character in the back, presenting an antinodal region between the transverse bars. The soundboard remains nearly motionless, except for a slight vibration in the lower zone, not visible in the experimental results. The analysis of the modal participation of the structural modes (see Fig. 6) reveals that the major contributor is mode $(1,3)b$, the participation of the rest of the modes being less than 10%. Regarding the acoustic modes, A0 and A1 participate in equal proportions, whereas A4 makes only a small contribution. Taking into account that the frequency of the $(1,3)b$ mode is nearly equal to the frequency of this mode—TB(8)—it is possible to infer the limited influence of the inside air on this modal parameter. Due to the slight vibration of the lower part of the soundboard, acting on the strings will excite this mode, although the strongest vibration will occur on the back.

IV. CONCLUSIONS

The finite-element and the modal analysis methods have been applied to the resonance box of a guitar with the purpose of analyzing its vibrational behavior. The numerical model was developed progressively, starting with the soundboard and back, then the assembled box and the inside air separately, and finally the whole box; that is, the wood structure and the air together. In this way, mode evolution can be tracked, establishing the influence of each component on the final box. Comparison of the modal patterns and frequencies with the modal analysis results corresponding to a real guitar box confirms the quality of the model. The newest feature is the successful development of the model to calculate and analyze the box–air coupling. In this sense we can conclude that

(1) Taking the air into account is essential to describe the vibrational modes of the guitar box;
(2) The soundboard–back coupling via the inside air is efficient in some cases, in which it generates modes of similar geometry, with in-phase or opposite phase vibrations;
(3) This coupling is decisive for the three lowest modes;
(4) Modes presenting a longitudinal flexural character in both the soundboard and back of the empty box are able to couple;
(5) Modes presenting a transverse flexural character either in the soundboard or in the back of the empty box produce noncoupled modes; that is, only one component vibrates;
(6) The Helmholtz resonance participates whenever soundboard and back couple;
(7) The influence of the inside air is decisive for the natural frequencies of the lowest modes of the box; and
(8) The Helmholtz resonance is dominant for the two lowest modes, whereas the participations of A0 and A1 tend to be similar for the higher coupled modes.

ACKNOWLEDGMENTS

The authors are grateful to Keller, S.A., for their important collaboration, and very much obliged to Alberto Atxotegi, the craftsman, for his careful work.

[1] E. V. Jansson, ''A study of the acoustical and hologram interferometric measurements of the top plate vibrations of a guitar,'' Acustica **25**, 95–100 (1971).
[2] K. A. Stetson, ''On modal coupling in string instrument bodies,'' J. Guitar Acoust. **3**, 23–29 (1981).
[3] I. M. Firth, ''Physics of the guitar at the Helmholtz and first top-plate resonances,'' J. Acoust. Soc. Am. **61**, 588–593 (1977).
[4] R. R. Boullosa, ''The use of transient excitation for guitar frequency response testing,'' Catgut Acoust. Soc. Newsletter **36**, 17–20 (1981).
[5] B. E. Richardson and G. W. Roberts, ''The adjustment of mode frequencies in guitars: A study by means of holographic interferometry and finite element analysis,'' SMAC'83, Vol. II (Publication of the Royal Swedish Academy of Music, No. 46:2), pp. 285–302 (1985).
[6] H. L. Schwab, ''Finite element analysis of a guitar soundboard I,'' Catgut Acoust. Soc. Newsletter **24**, 13–15 (1975).
[7] B. E. Richardson, G. W. Roberts, and G. P. Walker, ''Numerical model of two violins plates,'' J. Catgut Acoust. Soc. **47**, 12–16 (1987).
[8] J. Bretos, C. Santamaría, and J. Alonso-Moral, ''Vibrational patterns and frequency responses of the free plates and box of a violin obtained by finite element analysis,'' J. Acoust. Soc. Am. **105**, 1942–1950 (1999).
[9] K. D. Marshall, ''Modal analysis of a violin,'' J. Acoust. Soc. Am. **77**, 695–709 (1985).
[10] G. W. Caldersmith, ''Vibration geometry and radiation fields in acoustic guitars,'' Acoust. Australia **14**(2), 47–50 (1986).
[11] J. C. S. Lai and M. A. Burgess, ''Radiation efficiency of acoustic guitars,'' J. Acoust. Soc. Am. **88**, 1222–1227 (1990).
[12] M. J. Elejabarrieta, A. Ezcurra, and C. Santamaría, ''Evolution of the vibrational behavior of a guitar soundboard along successive construction phases by means of the modal analysis technique,'' J. Acoust. Soc. Am. **108**, 369–378 (2000).
[13] M. J. Elejabarrieta, A. Ezcurra, and C. Santamaría, ''Vibrational behavior of the guitar soundboard analyzed by the finite element method,'' Acust. Acta Acust. **87**, 128–136 (2001).
[14] ABAQUS (version 5.6), Hibbit, Karlsson, and Sorensen (1998).
[15] M. J. Elejabarrieta, A. Ezcurra, and C. Santamaría, ''Air cavity modes in the resonance box of the guitar: the effect of the sound hole,'' J. Sound Vib. (in press).
[16] SYSNOISE, Lms, Belgium (1994).
[17] D. J. Ewins, *Modal Testing: Theory and Practice* (Research Studies, Taunton, Somerset, England, 1984).
[18] E. V. Jansson, ''Acoustical properties of complex cavities and measurements of resonance properties of violin-shaped and guitar-shaped cavities,'' Acustica **37**(4), 211–221 (1977).
[19] T. D. Rossing, J. Popp, and D. Polstein, ''Acoustical response of guitars,'' SMAC'83, Vol. II (Publication of the Royal Swedish Academy of Music, No. 46:2), pp. 311–332 (1985).
[20] A. Runnemalm and N. E. Molin, ''Air cavity modes in sound boxes recorded by TV holography,'' *Proceedings of the 2nd Convention of the European Acoustics Association: Forum Acusticum*, Berlin (1999).
[21] B. E. Richardson, ''The acoustical development of the guitar,'' Catgut Acoust. Soc. J. **2**(5) (Series II), 1–10 (1994).
[22] G. Caldersmith, ''Plate fundamental coupling and its musical importance,'' Catgut Acoust. Soc. Newsletter **36**, 21–27 (1981).
[23] O. Christensen and B. B. Vistisen, ''Simple model for low-frequency guitar function,'' J. Acoust. Soc. Am. **68**, 758–766 (1980).
[24] F. T. Dickens, ''Analysis of the first and second vibration modes in a guitar using an èquivalent electrical circuit,'' Catgut Acoust. Soc. Newsletter **35**, 18–21 (1981).
[25] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer, New York, 1991).
[26] O. Christensen, ''Qualitative models for low frequency guitar function,'' J. Guitar Acoust. **6**, 10–25 (1982).

Efficiency, accuracy, and stability issues in discrete-time simulations of single reed wind instruments

Federico Avanzini[a)]
Dipartimento di Elettronica e Informatica, Università di Padova, Via Gradenigo 6/A, 35131 Padova, Italy

Davide Rocchesso[b)]
Dipartimento di Informatica, Università di Verona, Strada Le Grazie 15, 37134 Verona, Italy

(Received 1 June 2000; accepted for publication 11 February 2002)

A quantitative study of discrete-time simulations for a single reed physical model is presented. It is shown that when the continuous-time model is discretized, a delay-free path is generated in the computation. A general solution is proposed to this problem, that amounts to operating a geometrical transformation on the equations. The transformed equations are discretized using four different numerical methods. Stability properties of each method are assessed through analysis in the frequency domain. By comparing the discrete and continuous frequency responses, it is studied how the physical parameters are mapped by each method into the discrete-time domain. Time-domain simulations are developed by coupling the four digital reeds to an idealized bore model. Quantitative analysis of the simulations shows that the discrete-time systems produced by the four methods have significantly different behaviors, even when high sampling rates are used. As a result of this study, a general scheme for accurate and efficient time-domain simulations of the single reed model is proposed. [DOI: 10.1121/1.1467674]

PACS numbers: 43.75.Pq [ADP]

I. INTRODUCTION

Numerical simulations are commonly used by musical acousticians for investigating experimentally the functioning of single reed wind instruments.[1–3] A widely accepted approach divides the instrument into two main functional blocks, the acoustic bore (i.e., the resonator), and the reed–mouthpiece system (sometimes referred to as the *exciter*). The resonator can be described through its reflection function (see Schumacher[1]). In this case, the pressure wave $p^-(t)$ reflected from the bore is obtained as time convolution of the reflection function with the incoming pressure wave $p^+(t)$. Another technique, *waveguide modeling* (Smith[4]), divides the resonator into adjacent sections. In each section the wave propagation is simulated by means of two delay lines (left- and right-going). Terminations and junctions, as well as dissipation phenomena, are taken into account by inserting filtering elements in and between sections. Waveguide models are particularly accurate and efficient for simple cylindrical geometries, such as an idealized clarinet bore.

The airflow through the reed slit can be related to the reed opening and pressure through a nonlinear equation derived from the Bernoulli law.[5] In the simplest approximation, the reed is assumed to move in phase with the driving pressure and is described only in terms of its stiffness. This is sometimes referred to as the *quasistatic* approximation.[6] Using this approximation, the exciter can be described as a nonlinear map that relates flow and pressure at the bore entrance (see McIntyre *et al.*[2]). Despite its simplicity, such a quasistatic model is able to capture the basic nonlinear mechanisms that generate self-sustained oscillations in a single reed instrument. Due to its compactness and the low number of parameters, this reed model can be efficiently used for sound synthesis purposes.[4]

A more accurate model takes into account other mechanical properties of the reed, namely its mass and the dissipation due to internal losses and air friction. In a first approximation, these elements are incorporated in the model by describing the reed as a damped second-order oscillator.[1,5,7] This linear mechanical system is coupled with the nonlinear fluid dynamical equation. For clarity, in the following this model is referred to as the *dynamic* reed model.

In order to develop numerical simulations for the dynamic model, one has to overcome two main problems. First, the coupling in the equations typically generates a delay-free path in the computation. Due to the presence of a nonlinear equation, solving this loop is not trivial, unless one resorts to iterative solvers. Second, a numerical technique has to be chosen that preserves with reasonable accuracy the main properties of the physical system.

Concerning the first problem, many authors (see Gazengel *et al.*[3]) compute the discrete-time equations by inserting a fictitious delay in the delay-free path. However, the numerical error introduced in this way can lead to instability, especially at low sampling rates. Anderson and Spong[8] have proved analytically that the insertion of a delay element in a dynamical system deteriorates its stability properties. The second problem, i.e., the choice of an accurate discretization technique, is often neglected in the literature, and the equations are usually discretized using simple methods (such as the Euler method or the impulse invariant method), that introduce noticeable numerical artifacts in the simulations.

In this paper, an accurate and efficient discretization scheme for the dynamic reed model is developed. The delay-free loop in the computational scheme is solved by means of

a)Electronic mail: avanzini@dei.unipd.it
b)Electronic mail: rocchesso@sci.univr.it

TABLE I. Variables and constants in the reed model.

Quantity	Symbol	Value
Sampling period	T_s	(s)
Sampling rate	F_s	Hz)
Frequency of the continuous-time system	ω_c	(rad/s)
Frequency of the discretized systems	ω_d	(rad/s)
Reed opening	h	(m)
Reed equil. opening	h_0	$0.4\cdot 10^{-3}$ m
Reed displacement	y_r	(m)
Reed equil. displ.	y_0	(m)
Maximum allowed displ.	y_m	(m)
Reed mass/area	μ_r	0.0231 kg/m^2
Effective flow surface	S_r	$1.46\cdot 10^{-4}$ m^2
Reed resonance freq.	ω_r	23 250 rad/s
Reed damping	g_r	3 000 s^{-1}
Amplitude parameter	A	0.079 7 m^3/(N$^{2/3}$ s)
Mouth pressure	p_m	(Pa)
Mouthpiece pressure	p	(Pa)
Pressure drop	Δp	(Pa)
Mouthpiece vol. flow	u	(m^3/s)
Flow through the slit	u_f	(m^3/s)
Wave impedance of the bore	Z_0	2 290 133 kg/m^4 s
Speed of sound	c	347 m/s
Length of the bore	L	(m)
Bell cutoff freq.	f_{co}	$1\cdot 10^3$ Hz
Pr. wave from the bore	p^-	(Pa)
Pr. wave to the bore	p^+	(Pa)

the so-called *K method*, recently proposed by Borin *et al.*[9] The method operates a geometric transformation on the nonlinearity, in such a way that the delay-free path can be computed without introducing fictitious delays in the discrete-time equations.

Given a general solution to noncomputability problems, four different numerical methods are used for discretizing the mechanical differential equations. The so-obtained ''digital reeds'' are analyzed in the frequency domain and compared to the continuous-time system. Then, the digital reeds are connected to a waveguide model of an ideal cylindrical bore, and the resulting systems are compared in the time domain through numerical simulations.

The choice of the discretization method is usually considered to be noncritical when simulations are run at high sampling rates. However, the results presented here show that this choice does affect the behavior of the numerical models noticeably, even at a sampling rate of several tens of kHz (e.g., 44.1 kHz). Among the considered techniques, the 1-*step weighted sample* method is found to be the most suitable choice for discretizing the dynamic reed model. This method accurately preserves the properties of the physical system, and its low computational costs make it suitable for real-time applications.

Section II describes the dynamic reed model. Numerical techniques are discussed in Sec. III. The frequency analysis presented in Sec. IV compares the digital reeds with the continuous-time system. Finally, Sec. V discusses results obtained from time-domain simulations.

Notation and symbols. Table I summarizes the symbols used throughout the paper. The parameter values are taken from Schumacher.[1] Each time-varying variable x is written as $x(t)$ and $x(n)$, respectively, in the continuous and discrete-time domains. The transforms of the continuous and discrete-time signals are written, respectively, as $X(s)$ and $X(z)$.

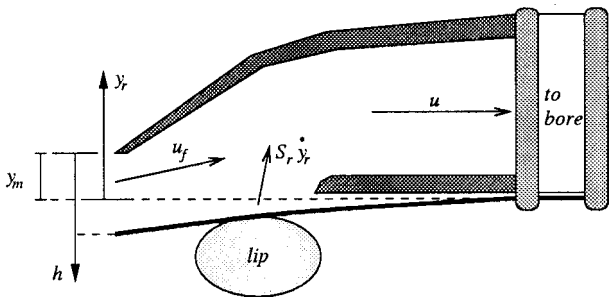


FIG. 1. Schematic representation of the reed–mouthpiece system.

II. THE PHYSICAL MODEL

A. Exciter

The dynamic model described below relies on the following assumptions: under normal playing conditions, oscillations occur mainly in the vertical direction, therefore a single degree of freedom (i.e., the reed tip vertical displacement) can be reasonably assumed; the reed resonances are well above the main frequency component of the driving pressure; therefore, only the effect of the first reed resonance needs to be modeled; the reed dimensions are small with respect to typical wavelengths in the resonator; thus, pressure can be thought of as constant along the reed internal surface. See Fig. 1.

Consequently, many authors[1,5,7] approximate the reed as a harmonic oscillator, driven by the pressure drop $\Delta p = p_m - p$ across the slit. When the reed beats (i.e., when it strikes the mouthpiece) an inelastic collision occurs:

$$\begin{aligned} &\ddot{y}_r(t) + g_r \dot{y}_r(t) + \omega_r^2 [y_r(t) - y_0] = \Delta p(t)/\mu_r, \\ &\qquad\qquad\qquad\qquad \text{if } y_r < y_m, \\ &y_r(t) = y_m \text{ and } \dot{y}_r(t) = 0, \quad \text{if } y_r \geqslant y_m. \end{aligned} \tag{1}$$

The total airflow u at the mouthpiece is the sum of two components. The first one is given by the flow u_f through the slit. The second component is produced by the reed motion and depends on the reed velocity. This component is assumed to be $S_r \dot{y}_r$, where S_r is the effective surface associated with the flow. Hence, the total flow u is given by

$$u(t) = u_f(t) + S_r \dot{y}_r(t). \tag{2}$$

The flow u_f through the slit is related to the pressure drop Δp and to the opening h via the nonlinear equation

$$\Delta p = F(u_f, h) := A^{-\alpha} \operatorname{sgn}(u_f) |u_f|^{\alpha} / h^2. \tag{3}$$

With the value $\alpha = 2$, Eq. (3) is the Bernoulli equation for an ideal fluid in the static regime. Using experiments on real instruments, Backus[5] found empirically a value for α different from the theoretical one, namely $\alpha = 3/2$. Backus suggested that this discrepancy could be due to the particular shape of the slit. More recently, Hirschberg *et al.*[10] questioned the validity of Backus' experiments. These authors developed a flow model that uses the standard Bernoulli equation (with $\alpha = 2$).

In order to account for air inertance the term $M_e\dot{u}_f$ must be added to the right-hand side of Eq. (3) (M_e being the effective mass through the slit). The inclusion of this term complicates the model, since the computation of $\dot{u}_f$ is required. According to many authors,[2,3,11] the effect of M_e is generally small and this additional term can be neglected.

In summary, the dynamic reed model adopted in this work is fully described by the set of equations (1), (2), (3).

B. Resonator

The acoustical bore can be described by means of pressure waves $p^{\pm}$, which by definition are related to pressure and airflow via the equations $p=p^+ + p^-$ and $u=(p^+ - p^-)/Z_0$. If cylindrical geometry is assumed and boundary losses are neglected, then the pressure wave p^+ coming from the mouthpiece propagates freely with speed c until it reaches the open end, terminated by the bell. The bell acts as a low-pass filter, reflecting low-frequency components inside the bore and radiating high-frequency components. Typical values for the cutoff frequency f_{co} of the bell are between 1 and 1.5 kHz. The pressure wave p^- reflected from the bore to the mouthpiece is thus given by

$$P^-(s) = -R_c(s)\exp(-s\ \ 2L/c)P^+(s). \tag{4}$$

$R_c(s)$ is the low-pass transfer function of the bell. The term $\exp(-s\ \ 2L/c)$ accounts for the delay $2L/c$ in the trip along the bore.

The waveguide model simulates the propagation with two delay lines. The length m_L of each line is chosen in such a way that $m_L c T_s = L$. In the discrete domain the delay term $\exp(-s2L/c)$ is therefore replaced by z^{-2m_L}. The low-pass bell filter is designed with standard techniques:[13] first an analog filter is designed using a fourth-order Butterworth realization; then, a digital equivalent $R_d(z)$ is obtained with usual methods from digital signal processing, such as the bilinear transformation

$$R_d(z) = R_c\left(2F_s\frac{1-z^{-1}}{1+z^{-1}}\right). \tag{5}$$

Summarizing, the digital bore model takes the incoming pressure wave p^+ from the exciter, and reflects it as an outgoing pressure wave p^- given by

$$P^-(z) = -R_d(z)z^{-2m_L}P^+(z). \tag{6}$$

III. NUMERICAL TECHNIQUES

The waveguide techniques outlined in Sec. II B provide a model for the resonator in the discrete time-space domain. This section discusses the discrete-time approximation of the dynamic reed model. Equations (1), (2), and (3) generate the system

$$\begin{aligned}\dot{\mathbf{w}}(t) &= \mathbf{A}\mathbf{w}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{c}\Delta p(t),\\ \mathbf{x}(t) &= \mathbf{D}\mathbf{w}(t) + \mathbf{E}\mathbf{u}(t) + \mathbf{f}\Delta p(t),\\ \Delta p(t) &= F(\mathbf{x}(t)),\end{aligned} \tag{7}$$

where

$$\mathbf{w}=\begin{bmatrix}h\\ \dot{h}\end{bmatrix},\quad \mathbf{u}=\begin{bmatrix}h_0\\ p_m\\ p^-\end{bmatrix},\quad \mathbf{x}=\begin{bmatrix}u_f\\ h\end{bmatrix}, \tag{8}$$

and

$$\mathbf{A}=\begin{bmatrix}0 & 1\\ -\omega_r^2 & -g_r\end{bmatrix},\quad \mathbf{B}=\begin{bmatrix}0 & 0 & 0\\ \omega_r^2 & 0 & 0\end{bmatrix},\quad \mathbf{c}=\begin{bmatrix}0\\ -1/\mu_r\end{bmatrix}, \tag{9}$$

$$\mathbf{D}=\begin{bmatrix}0 & -S_r\\ 1 & 0\end{bmatrix},\quad \mathbf{E}=\begin{bmatrix}0 & 1/Z_0 & -2/Z_0\\ 0 & 0 & 0\end{bmatrix},\quad \mathbf{f}=\begin{bmatrix}-1/Z_0\\ 0\end{bmatrix}. \tag{10}$$

The beating condition in Eq. (1) turns into

$$\mathbf{w}=\mathbf{0},\ \text{if}\ \ h\leqslant 0. \tag{11}$$

A. Solving the delay-free loop

When the first equation in system (7) is discretized, the structure of the resulting difference equation is found to be

$$\mathbf{w}(n) = \tilde{\mathbf{w}}(n) + \overline{\mathbf{c}}\Delta p(n). \tag{12}$$

The vector $\tilde{\mathbf{w}}(n)$ is a linear combination of all the terms that are computable at time n [namely, $\mathbf{u}(n)$ and past values of $\mathbf{w}$, $\mathbf{u}$, and Δp], while the vector $\overline{\mathbf{c}}$ weights the dependence of $\mathbf{w}$ on $\Delta p(n)$. Explicit expressions for both $\tilde{\mathbf{w}}(n)$ and $\overline{\mathbf{c}}$ depend on the numerical method actually used. The remaining equations in system (7) can thus be written as

$$\begin{aligned}\mathbf{x}(n) &= \tilde{\mathbf{x}}(n) + \mathbf{k}\Delta p(n),\\ \Delta p(n) &= F(\tilde{\mathbf{x}}(n) + \mathbf{k}\Delta p(n)),\end{aligned} \tag{13}$$

where $\mathbf{k}=(\mathbf{D}\overline{\mathbf{c}}+\mathbf{f})$. The vector $\tilde{\mathbf{x}}(n)=\mathbf{D}\tilde{\mathbf{w}}(n)+\mathbf{E}\mathbf{u}(n)$ does not depend on $\Delta p(n)$ and is therefore computable at each step. In the second equation (13), the term $\Delta p(n)$ depends implicitly on itself. In order to compute this equation, the *K method* is used.[9] This method uses the implicit mapping theorem to operate a geometric transformation on the nonlinear function F, and turns the second equation (13) into an explicit dependence

$$\begin{aligned}&\Delta p(n) = F(\tilde{\mathbf{x}}(n) + \mathbf{k}\Delta p(n)),\\ &\quad\overset{K\ \text{method}}{\mapsto}\quad \Delta p(n) = \overline{F}(\tilde{\mathbf{x}}(n)).\end{aligned} \tag{14}$$

Therefore, at each time-step n the vector $\tilde{\mathbf{x}}(n)$ is computed first, then $\Delta p(n)$ is obtained through the new nonlinear relation $\overline{F}$. Although $\overline{F}(\tilde{\mathbf{x}})$ is not available in closed form in most cases, an implementation can still be obtained without resorting to iterative solvers, by storing the multivariable function $\overline{F}$ as a set of precomputed tables.[14]

B. Discretizing the equations

The K method provides a robust and general means to compute the difference equations (13) accurately. Given such a method, different discretization techniques for system (7) can be compared.

Typical choices in the literature are the Euler method and the impulse invariant method. However, both these techniques introduce artifacts in the numerical systems. Gazengel *et al.*[3] discuss the use of a fourth-order Runge–Kutta solver.

This method, although very accurate when high sampling rates are used, turns out to be unstable at low F_s. Moreover, it has high computational costs, since four evaluations of the nonlinear function $F(\mathbf{x}(n))$ are needed at each time step. Van Walstijn[15] uses a hybrid backward-centered scheme that approximates the first derivative with the backward Euler rule and the second derivative with a centered difference. One advantage of this approach is that the vectors $\overline{\mathbf{c}}$ and $\mathbf{k}$ in Eqs. (12) and (13) are both zero; therefore, no delay-free paths are created in the discrete-time equations. However, at each time step n the Newton–Raphson method is used for computing iteratively the flow $u(n)$, and nine iterations are typically required.

The following techniques are used in the rest of this paper.

(i) *1- and 2-step Adams-Moulton methods* (AM1, 2 from now on). These are linear multistep methods, whose stability and accuracy properties are known from the numerical analysis literature.[16]

(ii) *1- and 2-step weighted sample methods* (WS1, 2 from now on). These have been introduced recently by Wan and Schneider.[17] They are designed for generic linear systems, and are based on a polynomial interpolation of the input.

Higher-order methods are not used for two main reasons: (1) stability properties tend to deteriorate with increasing order, and (2) the computational costs become higher.

As Schneider *et al.*[18] have pointed out, AM methods can be easily seen as s-to-z mappings in the complex plane

$$(\mathrm{AM1}) \quad s = 2F_s \frac{1-z^{-1}}{1+z^{-1}}, \tag{15a}$$

$$(\mathrm{AM2}) \quad s = 12F_s \frac{1-z^{-1}}{5+8z^{-1}-z^{-2}}. \tag{15b}$$

Note that the mapping (15a) associated to the AM1 method is the bilinear transformation.

Applying the AM methods to the first equation in system (7) amounts to Laplace-transforming it and substituting each occurrence of s with the corresponding mapping (15a) or (15b). Therefore, the differential equation is turned into a second-order and a fourth-order difference equation, by the AM1 and the AM2 methods, respectively. If system (7) is time-invariant, then the AM methods provide a time-invariant discrete system. If some of the reed parameters are varied over time, then the discrete system coefficients need to be updated at a suitable control rate.

Wan and Schneider[17] have shown that the k-step WS method turns the first equation in system (7) into the difference equation

$$\mathbf{w}(n) = \mathbf{\Phi}(T_s)\mathbf{w}(n-1) + \boldsymbol{\mathcal{W}}_{\mathbf{u}}\mathbf{S}_{\mathbf{u}}^{(k)} \begin{bmatrix} \mathbf{u}(n) \\ \vdots \\ \mathbf{u}(n-k) \end{bmatrix} + \boldsymbol{\mathcal{W}}_{\Delta p}\mathbf{S}_{\Delta p}^{(k)} \begin{bmatrix} \Delta p(n) \\ \vdots \\ \Delta p(n-k) \end{bmatrix}, \tag{16}$$

where $k=1,2$. Therefore, the differential equation is turned into a second-order and a third-order difference equation, respectively, by the WS1 and the WS2 methods.

Details about the computation of the matrices can be found elsewhere.[17,19] The matrices $\mathbf{S}_{\mathbf{u}}^{(k)}$ and $\mathbf{S}_{\Delta p}^{(k)}$ are interpolation matrices that depend only on the order k of the method, while $\mathbf{\Phi}(T_s)$ is the exponential matrix defined from $\mathbf{\Phi}(t)=\exp(\mathbf{A}t)$. Computation of $\boldsymbol{\mathcal{W}}_{\mathbf{u}}$, $\boldsymbol{\mathcal{W}}_{\Delta p}$ involves calculation of the $k+1$ integrals $\int_0^T \mathbf{\Phi}(T_s-t)\cdot t^l\,dt$ (for $l=1\cdots k+1$). Therefore, computation of the coefficients in the difference equation (16) requires computation of transcendental functions.

If system (7) is time-invariant, then computation of matrices $\mathbf{\Phi}(T_s)$, $\boldsymbol{\mathcal{W}}_{\mathbf{u}}$, and $\boldsymbol{\mathcal{W}}_{\Delta p}$ can be performed off-line, while these matrices need to be updated at control rate when time-varying control parameters are used. In this latter case the WS methods have higher computational costs than the AM methods, and this is a potential drawback for real-time applications. However, in the case of low-order methods ($k=1,2$), only a small number of coefficients needs to be updated. Moreover, Wan and Schneider[17] show that the computational costs can be lowered using *ad hoc* techniques (e.g., the columns of $\boldsymbol{\mathcal{W}}$ can be computed iteratively).

C. Properties of the methods

Stability properties of an AM method are summarized by the shape of its region of absolute stability $\mathcal{R}_A$ (see Lambert).[16] If the continuous-time eigenvalues λ_l ($l=1,...,N$) of a stable, N-dimensional continuous system lie inside $\mathcal{R}_A$, then the discretized system is stable. As far as the bilinear transformation is concerned, it is known that $\mathcal{R}_A$ is the whole left-half s-plane. Therefore, continuous-time eigenvalues λ_l with $\mathrm{Re}(\lambda_l)<0$ are mapped into discrete-time eigenvalues p_l with $|p_l|<1$, and stability is preserved at any sampling rate. The AM2 method has worse stability properties, since its region of absolute stability is the finite subset of the left-half s-plane shown in Fig. 2. This means that stability is preserved only at high sampling rates, so that the eigenvalues of the continuous system lie inside $\mathcal{R}_A$.

For the k-step WS methods, Wan and Schneider have shown that the discrete-time eigenvalues are the roots of the characteristic polynomial $p(z)=z^{k-1}|z\mathbf{I}-\mathbf{\Phi}(T_s)|$. Therefore, continuous-time eigenvalues λ_l map into discrete-time eigenvalues p_l through the relation

$$p_l = \exp(\lambda_l T_s). \tag{17}$$

If $\mathrm{Re}(\lambda_l)<0$, then $|p_l|<1$; therefore, stability is preserved at any sampling rate. Note that the same relation (17) between discrete- and continuous-time eigenvalues holds for the impulse invariant method. Indeed, it can be verified that the WS method with $k=0$ is completely equivalent to the impulse invariant method.

Concerning accuracy, it is a general result that the k-step AM method has order $k+1$. This means that the methods provide a global truncation error in time which has order T_s^{k+1}. For the WS methods, Wan and Schneider have given experimental results showing that a k-step method has order $k+1$, the same as the corresponding AM method.

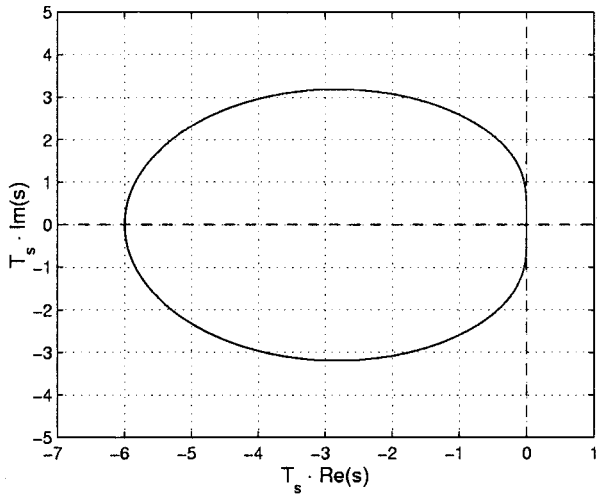


FIG. 2. Region of absolute stability $\mathcal{R}_A$ for the AM2 method.

IV. THE DIGITAL REEDS

This section is devoted to frequency analysis of the digital reeds obtained using AM and WS methods. Such analysis permits comparison of the different discretization techniques, by studying how the physical parameters are mapped into the discrete-time domain.

Consider the transfer function $H_c(s)$ of the relative displacement $y_r - y_0$ versus the pressure drop Δp. From Eq. (1), this is seen to be the harmonic oscillator

$$H_c(s) = \frac{1}{\mu_r} \frac{1}{s^2 + g_r s + \omega_r^2}. \tag{18}$$

The frequency response is given by $H_c(j\omega_c)$. A meaningful comparison between the numerical methods amounts to analyzing how they preserve this frequency response in the discrete-time domain. The study is performed for various F_s. Typical audio sampling rates $F_s = 22.05$ kHz and $F_s = 44.1$ kHz are taken as reference values. Following ideas developed by Gazengel *et al.*,[3] the analysis is focused on three physically meaningful parameters of the dynamic model: the resonance frequency ω_r, the oscillator stiffness (defined as $|H_c(0)|^{-1} = \mu_r \omega_r^2$), and the damping coefficient g_r.

Typical values for ω_r lie in the high-frequency region, and this parameter is therefore considered to be noncritical in helping self-sustained oscillations. Indeed, self-sustained oscillations occur even when there is no resonance at all, as in the quasistatic approximation (see Sec. V B in the following). However, as pointed out by Thompson,[12] the reed resonance has a role in adjusting pitch, loudness, and tone color, as well as helping transitions to high regimes of oscillation, such as the clarion register or the reed regime (''squeaks''). Stiffness characterizes the reed response in the low-frequency region and is therefore an important parameter, since the fundamental frequency of the oscillation always lies in this region. Concerning the damping coefficient, the relation $g_r = \omega_r / Q_r$ holds for the harmonic oscillator (18), where $Q_r = \omega_r / (\omega_1 - \omega_2)$ is the quality factor and $\omega_{1,2}$ are the 3-dB cutoff frequencies. Therefore, $g_r = \omega_1 - \omega_2$ for the continuous-time oscillator (18).

When using AM and WS methods, $H_c(s)$ is turned into a digital filter which is not a harmonic oscillator. Therefore, the parameters ω_r, μ_r, g_r cannot be deduced from the coefficients of the discrete-time transfer functions. Instead, they are extrapolated from the shape of the discrete-time frequency responses. In particular, following Gazengel *et al.*[3] the digital damping coefficient is defined in the following as $g_d = \omega_{d1} - \omega_{d2}$, where $\omega_{d1,d2}$ are the 3-dB cutoff frequencies for the response of the discrete-time system.

A. Adams–Moulton methods

Using the AM methods, the digital transfer functions $H_{\mathrm{AM1}}(z)$ and $H_{\mathrm{AM2}}(z)$ are obtained by substitution of the corresponding s-to-z mapping [(15a), (15b)] in $H_c(s)$. The corresponding frequency responses are given by evaluation at $z = \exp(j\omega_d / F_s)$. From Sec. IV, $H_{\mathrm{AM1}}(z)$ and $H_{\mathrm{AM2}}(z)$ are known to have order 2 and 4, respectively.

The frequency responses H_{AM1} and H_{AM2} of the discretized systems are plotted in Fig. 3 for the cases $F_s = 22.05$ kHz and $F_s = 44.1$ kHz. Responses obtained with the Euler method are also plotted as a term of comparison.

The Euler method is seen to provide poor accuracy even at $F_s = 44.1$ kHz. In particular, a noticeable numerical dissipation is introduced, so that the resonance is heavily attenuated. Results for AM methods are in good agreement with theoretical predictions. Both the magnitude and the phase responses of H_{AM1} exhibit a known ''frequency warping'' phenomenon: the induced map between the continuous frequency ω_c and discrete frequency ω_d is (see Mitra)[13] $\omega_d = 2F_s \arctan(\omega_c / 2F_s)$. High frequencies are thereby compressed, and this phenomenon becomes more noticeable as the sampling rate decreases; at $F_s = 22.05$ kHz the resonance ω_r of H_{AM1} has shifted from the original value 23 250 rad/s to the value 21 300 rad/s (i.e., from 3700 Hz to 3390 Hz).

The AM2 method provides different results: there is no significant warping, but the magnitude of the resonance is amplified. The amplification is small at $F_s = 44.1$ kHz, but becomes unacceptable at $F_s = 22.05$ kHz (the peak magnitude is $4.7 \cdot 10^{-5}$ m/Pa). This phenomenon is a direct consequence of stability properties. Indeed, it can be seen that, using the values listed in Table I, the method becomes unstable at $F_s \simeq 19$ kHz. This explains the strong amplification and the phase distortion exhibited by H_{AM2} at $F_s = 22.05$ kHz.

Both the methods preserve stiffness [the equalities $H_{\mathrm{AM1}}(0) = H_{\mathrm{AM2}}(0) = H_c(0)$ hold]. Finally, qualitative analysis shows that both methods lead to a digital damping $g_d = \omega_{d1} - \omega_{d2}$ that is smaller than the physical one, and decreases with decreasing F_s. For H_{AM1} this is a consequence of frequency warping, which causes the resonance bandwidth to reduce, and the quality factor to increase consequently. For H_{AM2} this is due to the resonance amplification rather than to warping.

B. Weighted sample methods

The 1- and 2-step WS methods do not define a s-to-z mapping; therefore, the discrete-time transfer functions

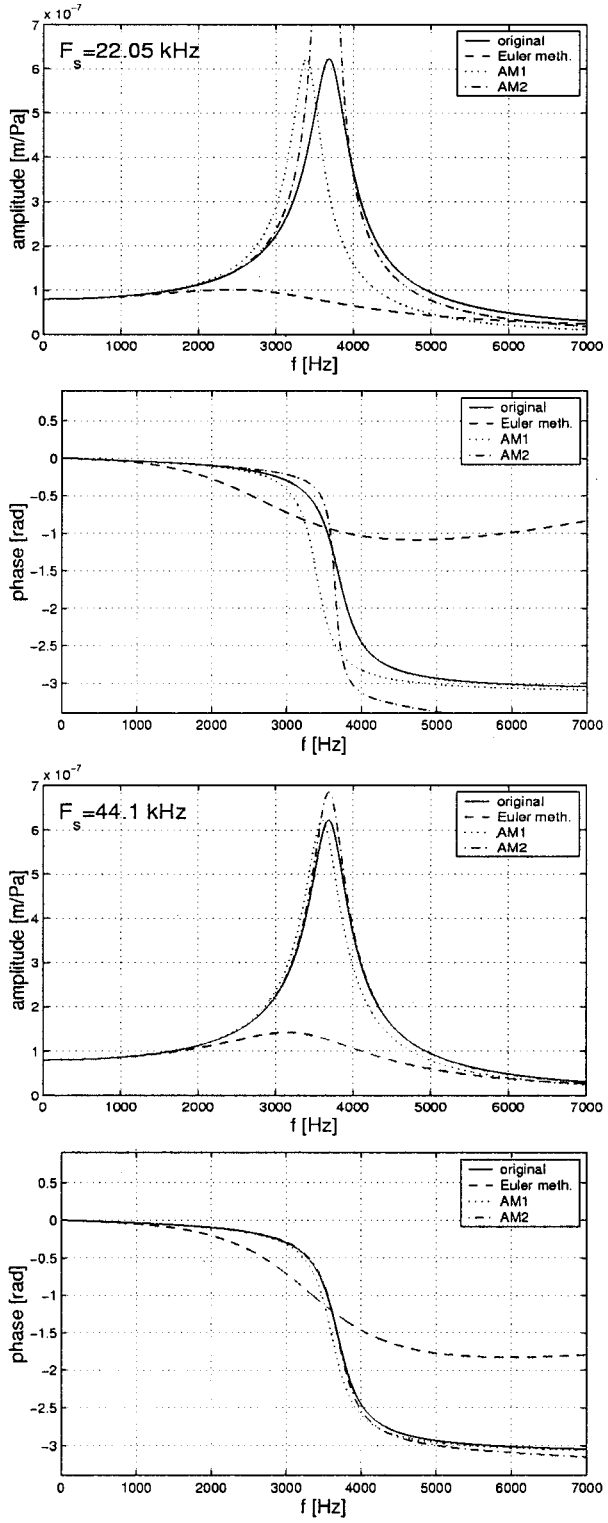


FIG. 3. Frequency responses from AM methods, with $F_s = 22.05, 44.1$ kHz. H_c (solid line), Euler method discretization (dashed), H_{AM1} (dotted), H_{AM2} (dot-dashed).

$H_{\mathrm{WS1}}(z)$ and $H_{\mathrm{WS2}}(z)$ are not obtained by substitution. Instead, they are computed directly from the general equation (16). From Sec. IV, it can be seen that $H_{\mathrm{WS1}}(z)$ and $H_{\mathrm{WS2}}(z)$ have order 2 and 3, respectively.

Results are summarized in Fig. 4. Responses obtained with the Euler method are plotted as a term of comparison.

The discrete-time responses H_{WS1} and H_{WS2} show excellent agreement with H_c, even at low sampling rates. Both methods preserve the resonance ω_r without introducing warping. Stiffness is preserved as well. Numerical dissipation is introduced, which is more significant for the 1-step method. This can be noticed by observing that the digital amplitude responses lie below the continuous one. Due to this dissipation, the digital damping coefficient g_d is larger than the physical one and increases with decreasing F_s, for both $H_{\mathrm{WS1}}(z)$ and $H_{\mathrm{WS2}}(z)$. Phase responses are well preserved by both methods.

Summarizing, the frequency analysis developed in this section has shown that the WS methods better approximate the reed frequency response than AM methods. It would appear that the WS methods are preferable. However, this conjecture is not confirmed from the time-domain analysis developed in the next section.

V. TIME-DOMAIN SIMULATIONS

In order to obtain time-domain simulations, each of the four digital reeds is coupled to the same resonator (the waveguide cylindrical bore described in Sec. II) to form a complete instrument. Comparisons of simulations lead to two main results, which are not evident from the frequency analysis of Sec. IV.

(i) the systems can behave differently even at high sampling rates, where the reed frequency responses are indistinguishable;
(ii) both the 1-step methods approximate the continuous system accurately, while the 2-step methods exhibit artifacts.

A. Threshold pressure

A first simulation study concerns the threshold pressure p_t, defined as the value of mouth pressure above which stable oscillations take place. A rough estimate for the threshold pressure, $p_t \simeq h_0 \mu_r \omega_r^2/3$, can be derived using the *quasistatic* approximation (see Hirschberg *et al.*[6] and Sec. V B). With the values listed in Table I, the ''quasistatic estimate'' is $p_t \simeq 1664$ Pa. However, as observed by Keefe,[20] this value underestimates the true p_t.

In this section such a quasistatic estimate is compared with experimental results from simulations. First, a ''dynamic estimate'' is found by running simulations at very high sampling rates (up to 500 kHz). For such sampling rates, all the systems are found to have the same threshold pressure, $p_t = 1802$ Pa. This is therefore assumed to be the ''true'' value. Then, simulations are run at lower sampling rates: the measured p_t are given in Table II, from which some remarks can be made.

(i) For all the digital reeds, p_t converges to the dynamic estimate 1802 Pa as the sampling rate is increased. The convergence of AM2 is not evident from Table II, since it occurs at $F_s > 200$ kHz.
(ii) The p_t estimates obtained from both the 1-step methods exhibit robustness with respect to the sampling rate. At $F_s = 30$ kHz, the deviation of p_t from the true value is less than 1% for both AM1 and WS1.

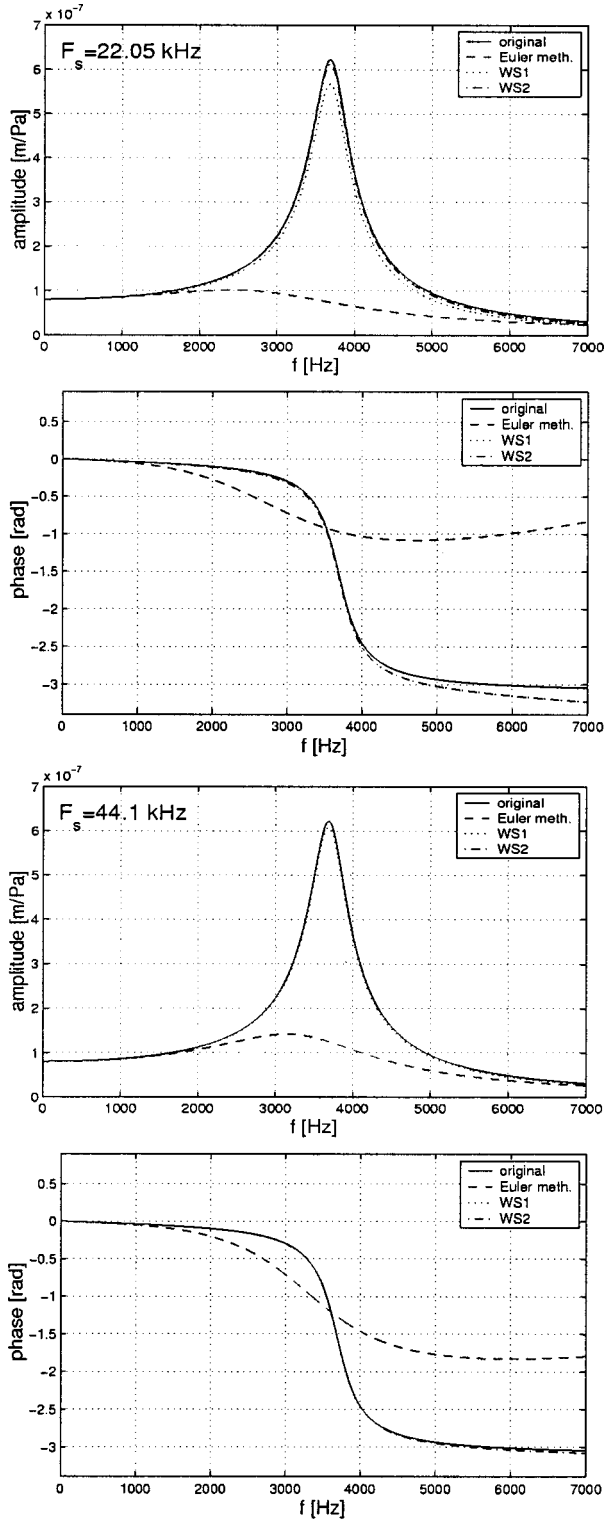


FIG. 4. Frequency responses from WS methods, with $F_s = 22.05, 44.1$ kHz. H_c (solid line), Euler method discretization (dashed), H_{WS1} (dotted), H_{WS2} (dot-dashed).

(iii) The 2-step methods are less robust: even at high sampling frequencies, the threshold pressures for the corresponding systems are far from the true value. In particular, simulations with AM2 hardly reach steady-state oscillations for $F_s < 35$ kHz. For this reason the AM2 column in Table II provides results only for $F_s > 50$ kHz.

Consider the frequency response H_{WS2} at $F_s = 44.1$ kHz: from Fig. 4, this is seen to be indistinguishable from the original one. However, the estimated p_t is still noticeably higher than the true value. An analogous remark holds for AM2 with $F_s = 100$ kHz. Therefore, the 2-step methods exhibit poor accuracy even when the reed response is well approximated.

These results show that the discrete-time frequency response does not provide sufficient information on the digital reed when this is coupled with the nonlinearity $F(u_f, h)$ and with the bore. Due to the nonlinearity, the whole system exhibits sensitive dependence on small deviations in the frequency response.

B. Dynamic versus quasistatic

In this section, the dynamic model is compared with the *quasistatic* reed approximation. This approximation provides a simplified description of reed motion by exploiting the following: the reed resonance is noticeably higher than the playing frequency of any playable note on the clarinet; therefore, the spectrum of Δp is confined to the low-frequency region, where it is reasonable to assume a flat reed response (see Figs. 3, 4). The response $H_c(j\omega_c)$ is therefore approximated by the zero-frequency response $H_c(0) = 1/(\mu_r \omega_r^2)$, and in this approximation the reed moves in phase with the pressure drop according to the relation

$$[y_r(t) - y_0] = H_c(0)\Delta p(t). \tag{19}$$

Substituting Eq. (19) in Eq. (3), and recalling that $h = y_m - y_r$, a few calculation steps yield

$$u_f = A(h_0 - \Delta p/(\mu_r \omega_r^3))^{4/3} \cdot \mathrm{sgn}(\Delta p)|\Delta p|^{2/3}. \tag{20}$$

The properties of this memoryless model have been studied by many authors.[2,4,6,21]

A first comparison between the quasistatic and the dynamic models amounts to plotting the corresponding phase diagrams for the steady-state signals u_f and Δp. Figure 5 shows an example of such phase diagram, obtained from numerical simulations with the WS1 method with a mouth pressure $p_m = 2265$ Pa. This value is the maximum value for nonbeating conditions.

The phase diagram for the quasistatic model is simply the plot of Eq. (20), while the dynamic model exhibits a more interesting behavior: u_f and Δp move along a hysteretic path. This is due to the presence of memory in the equations: when the reed dynamics is taken into account, then Eqs. (19), (20) do not hold, and h and u_f depend on Δp together with its derivatives. In other words, the attractor in the dynamic reed model is not a curve in the plane, but instead a closed orbit in a higher-dimensional phase space.

Another important difference in the behavior of quasistatic and dynamic simulations concerns transitions to high regimes of oscillation. As Thompson[12] and Wilson and Beavers[7] have pointed out, both ω_r and g_r play a key role in helping transition to the second register (clarion register). Experiments with artificial lips and real clarinets have shown that the clarion register can be produced without opening the register hole, if the reed resonance matches a low harmonic of the playing frequency and the damping is small enough.

TABLE II. Measured threshold pressures from time-domain simulations.

	p_t (Pa)				
F_s (kHz)	Q. Static	AM1	WS1	AM2	WS2
20	1664	1816	1761	⋯	3346
25	1664	1808	1774	⋯	2842
30	1664	1807	1784	⋯	2554
35	1664	1807.5	1790	⋯	2365
40	1664	1807.5	1795	⋯	2233
45	1664	1804	1796	⋯	2136
50	1664	1804.5	1797	3781	2063
55	1664	1805	1798	3516	2008
60	1664	1805.5	1799	3278	1960
65	1664	1806	1799.5	3148	1932
70	1664	1803.5	1800	3026	1906
75	1664	1804	1800.5	2908	1881
80	1664	1804.5	1801	2841	1865
85	1664	1805	1801.5	2887	1848
90	1664	1803	1802	2737	1832
95	1664	1803.5	1802	2692	1816
100	1664	1803.5	1802	2643	1802

Moreover, extremely low damping causes the reed regime (''squeaks'') to be produced. From a musical standpoint, squeaks are often explained as a consequence of insufficient breathing, while the fundamental register comes in as mouth pressure is increased.

All these effects are reproduced using the dynamic model, while the quasistatic model does not provide control over such effects. Figure 6(a) shows examples of transitions from numerical simulations with the WS1 method. The clarion register is obtained by matching ω_r to the seventh harmonic of the playing frequency and lowering g_r to 1400 s^{-1}. In Fig. 6(b) the reed regime is achieved by giving g_r a value as low as 300 s^{-1}. Squeaks are more easily obtained in simulations by driving the reed with low blowing pressures.

VI. CONCLUSIONS

In this paper, a simulation scheme for the single reed system was proposed, resulting in two conclusions. First, analysis of the discrete-time frequency responses was performed in order to study how theoretical properties of the numerical methods, such as stability, accuracy, and frequency warping, affect the reed behavior in the digital domain. It has been shown that 1-step methods, such as the bilinear transformation or the 1-step weighted sample method, can approximate the system with good accuracy while keeping the computational costs low. Second, time-domain simulations were obtained by coupling the single reed to a simple bore. These have shown that the analysis of the reed frequency responses do not provide enough information on the properties of the whole system due to nonlinear coupling in the equations. Indeed, the discrete-time models have significantly different behaviors even when the

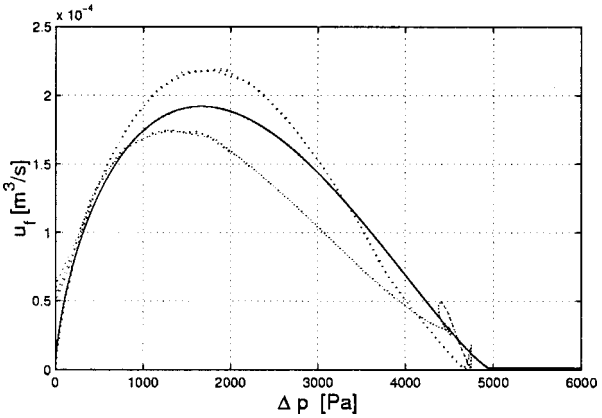


FIG. 5. Phase diagrams for u_f vs Δp. Quasistatic model (solid line), and dynamic model (dotted line) discretized with the WS1 method ($F_s = 44.1$ kHz, $p_m = 2265$ Pa).

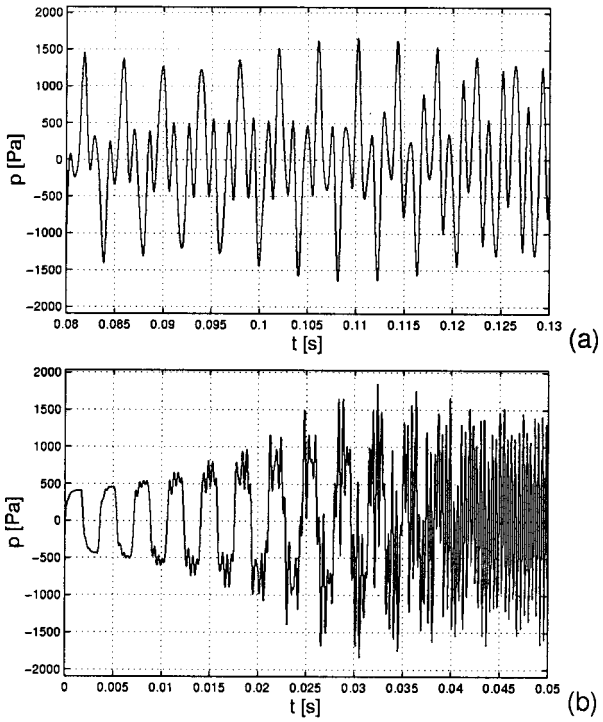


FIG. 6. Transitions to high regimes of oscillation (WS1 method, $F_s = 44.1$ kHz); (a) clarion register ($\omega_r = 2\pi \cdot 2020$ rad/s, $g_r = 1400\ s^{-1}$, $p_m = 1800$ Pa); (b) reed regime ($\omega_r = 2\pi \cdot 3150$ rad/s, $g_r = 300\ s^{-1}$, $p_m = 1900$ Pa).

frequency responses of the digital reeds are almost indistinguishable. Furthermore, the 2-step methods introduce artifacts in the digital domain while 1-step methods better approximate the original system.

These two results show that the 1-step weighted sample method provides the most accurate reed simulations. Due to its low computational costs, it can be used as an efficient tool for sound synthesis purposes, in combination with the K method. Moreover, low sampling rates can be used without introducing instability or serious artifacts. This is important in emerging applications such as structured audio coding,[22] where instrument models are encoded and transmitted without precise knowledge of the computational power of the decoder.

The present study has focused on the accurate simulation of the instrument exciter. Good sound quality from numerical simulations can only be achieved if also the resonator is modeled accurately (dissipation and dispersion, toneholes, radiation from the bell, etc.);[23] therefore, the bore model presented in Sec. II B is far too poor to provide satisfactory sound quality. Nonetheless, sound examples obtained from numerical simulations[24] show that reed physical parameters allow effective and realistic control over the digital instrument.

ACKNOWLEDGMENTS

The authors would like to thank Gianpaolo Borin and Giovanni De Poli for constant support throughout this research, and Mark Kahrs for providing a number of helpful remarks and comments on a preliminary version of the manuscript.

[1] R. T. Schumacher, ''*Ab initio* calculations of the oscillations of a clarinet,'' Acustica **48**(2), 71–85 (1981).
[2] M. E. McIntyre, R. T. Schumacher, and J. Woodhouse, ''On the oscillations of musical instruments,'' J. Acoust. Soc. Am. **74**, 1325–1345 (1983).
[3] B. Gazengel, J. Gilbert, and N. Amir, ''Time domain simulation of single reed wind instrument. From the measured impedance to the synthesis signal. Where are the traps?,'' Acta Acust. **3**, 445–472 (1995).
[4] J. O. Smith III, ''Principles of Digital Waveguide Models of Musical Instruments,'' in *Applications of DSP to Audio and Acoustics*, edited by M. Kahrs and K. Brandenburg (Kluwer Academic, Dordrecht, 1998), Chap. 10, pp. 417–466.
[5] J. Backus, ''Small-vibration theory of the clarinet,'' J. Acoust. Soc. Am. **35**, 305–313 (1963).
[6] A. Hirschberg, X. Pelorson, and J. Gilbert, ''Aeroacoustics of musical instruments,'' Meccanica **31**, 131–141 (1996).
[7] T. A. Wilson, and G. S. Beavers, ''Operating modes of the clarinet,'' J. Acoust. Soc. Am. **56**, 653–658 (1974).
[8] R. J. Anderson and M. W. Spong, ''Bilateral control of teleoperators with time delay,'' IEEE Trans. Autom. Control **34**, 494–501 (1989).
[9] G. Borin, G. De Poli, and D. Rocchesso, ''Elimination of delay-free loops in discrete-time models of nonlinear acoustic systems,'' IEEE Trans. Speech Audio Process. **8**(5), 597–606 (2000).
[10] A. Hirschberg, R. W. A. van de Laar, J. P. Marrou-Mauriéres, A. P. J. Wijnands, H. J. Dane, S. G. Kruijswijk, and A. J. M. Houtsma, ''A quasi-stationary model of air flow in the reed channel of single-reed woodwind instruments,'' Acustica **70**, 146–154 (1990).
[11] C. J. Nederveen, *Acoustical Aspects of Woodwind Instruments* (F. Knuf, Amsterdam, 1969).
[12] S. C. Thompson, ''The effect of the reed resonance on woodwind tone production,'' J. Acoust. Soc. Am. **66**, 1299–1307 (1979).
[13] S. K. Mitra, *Digital Signal Processing. A Computer Based Approach* (McGraw-Hill, New York, 1998).
[14] D. Rocchesso and F. Turra, ''A generalized excitation for real-time sound synthesis by physical models,'' in Proc. Stockholm Mus. Acoust. Conf., Stockholm (1993), pp. 584–588.
[15] M. van Walstijn, ''Discrete-Time Modelling of Brass and Reed Woodwind Instruments with Applications to Musical Sound Synthesis,'' Ph.D. dissertation, Dept. of Physics and Astronomy, University of Edinburgh, 2001.
[16] J. D. Lambert, *Numerical Methods for Ordinary Differential Systems* (Wiley, New York, 1993).
[17] C. Wan and A. M. Schneider, ''Further improvements in digitizing continuous-time filters,'' IEEE Trans. Signal Process. **45**(3), 533–542 (1997).
[18] A. M. Schneider, J. T. Kaneshige, and F. D. Groutage, ''Higher order s-to-z mapping functions and their application in digitizing continuous-time filters,'' Proc. IEEE **79**(11), 1661–1674 (1991).
[19] F. Avanzini, ''Stable and Accurate Numerical Methods in Physical Models of Musical Instruments. A Study on the Clarinet,'' in Proceedings of the Diderot Forum on Mathematics and Music, Vienna (1999), pp. 11–19.
[20] D. H. Keefe, ''On Sound Production in Reed-driven Wind Instruments,'' Systematic Musicology Technical Report 9003, School of Music, University of Washington, Seattle (1990).
[21] C. Maganza, R. Caussé, and L. Laloë, ''Bifurcations, period doublings and chaos in clarinetlike systems,'' Europhys. Lett. **1**(6), 295–302 (1986).
[22] E. D. Scheirer, ''Structured audio and effects processing in the MPEG-4 multimedia standard,'' Multimedia Syst. **7**(1), 11–22 (1999).
[23] G. P. Scavone, ''An Acoustic Analysis of Single-Reed Woodwind Instruments with an Emphasis on Design and Performance Issues and Digital Waveguide Modeling Techniques,'' Ph.D. dissertation, Dept. of Music, Stanford University, 1997.
[24] See ⟨www.dei.unipd.it/~avanzini/jasa2002/⟩.

Acoustic competition in the gulf toadfish *Opsanus beta*: Acoustic tagging

Robert F. Thorson[a] and Michael L. Fine[b]
Department of Biology, Virginia Commonwealth University, Richmond, Virginia 23284-2012

(Received 28 September 2001; revised 9 January 2002; accepted 6 February 2002)

Nesting male gulf toadfish *Opsanus beta* produce a boatwhistle advertisement call used in male–male competition and to attract females and an agonistic grunt call. The grunt is a short-duration pulsatile call, and the boatwhistle is a complex call typically consisting of zero to three introductory grunts, a long tonal boop note, and zero to three shorter boops. The beginning of the boop note is also gruntlike. Anomalous boatwhistles contain a short-duration grunt embedded in the tonal portion of the boop or between an introductory grunt and the boop. Embedded grunts have sound-pressure levels and frequency spectra that correspond with those of recognized neighbors, suggesting that one fish is grunting during another's call, a phenomenon here termed acoustic tagging. Snaps of nearby pistol shrimp may also be tagged, and chains of tags involving more than two fish occur. The stimulus to tag is a relatively intense sound with a rapid rise time, and tags are generally produced within 100 ms of a trigger stimulus. Time between the trigger and the tag decreases with increased trigger amplitude. Tagging is distinct from increased calling in response to natural calls or stimulatory playbacks since calls rarely overlap other calls or playbacks. Tagging is not generally reciprocal between fish, suggesting parallels to dominance displays. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1466865]

PACS numbers: 43.80.Ka, 43.80.Lb [WA]

I. INTRODUCTION

Toadfish produce sounds by contracting extremely fast muscles on the sides of a heart-shaped swimbladder.[1–3] As originally shown by Skoglund,[3] the muscle contraction rate generates the fundamental frequency of their sounds. Toadfish of both sexes in the subfamily *Batrachoidinae* produce short pulsatile grunts in agonistic situations,[4–7] and nesting males produce a long-tonal boatwhistle advertisement call[6,8,9] that functions in male–male competition and in female choice.[5,10–12] Males increase their calling rate in response to calls or playbacks of nearby males,[10–12] and females have been attracted to playbacks in pen tests.[12] The boatwhistle begins with a gruntlike component before exhibiting a clear fundamental frequency with harmonics. *Opsanus tau* and *Haplobatrachus didactylus* produce a boatwhistle of a single note,[8,9] but the gulf toadfish *Opsanus beta* produces a more complex call.[6,13] Its boatwhistle [Fig. 3(a)] includes from zero to three introductory grunt pulses followed by a long tonal boop note and up to three shorter boops.[13] Males call occasionally and irregularly during the day and increase their calling rates around sunset.[13,14] Calling males remain in nests for extended periods,[5,12] which has allowed us to separate the calls of individual fish by their sound parameters.[13]

Fine[15] recorded anomalous boatwhistles of *Opsanus tau* that contained a grunt embedded within the tonal portion of the call. We examined this phenomenon in the gulf toadfish *Opsanus beta* living in a canal in the Florida Keys and present evidence that it is caused by a second fish grunting during the call, i.e., acoustic tagging.

[a]Present address: 133 Mockingbird Rd., Tavernier, FL 33070.
[b]Electronic mail: mfine@atlas.vcu.edu

II. METHODS

Boatwhistles were recorded from a hydrophone installation placed on the bottom of a bayside canal at Plantation Key, Islamorada, FL. We listened to over 300 hours of underwater sound in 2000 (sounds were heard from January 9 to April 4) and over 200 hours in 2001 (28 December 2000 to 4 May 2001). Most attention was devoted to the crepuscular period, when calling rate increases, although recordings were made at various times. We present detailed data (quantification of all grunts and boatwhistles) from a typical 2-h twilight recording from 12 March 2000. Additional qualitative and quantitative observations from other recordings indicate these results are typical until near the end of the mating season (see Sec. III).

The hydrophone was at approximately 3 m depth and about 3 m from the sea wall. The canal is about 26 m wide and 230 m long and leads out to Florida Bay. The sea wall is made of coquina block and extends down about 2 m. The bottom of the canal is coral covered by silt and vegetation. The hydrophone installation remained fixed in the water during the course of the study. Since *Opsanus tau* boatwhistles attenuate rapidly with distance in shallow water,[16] we tested the hypothesis that acoustic interactions and tagging are restricted to nearby fish by making several additional stereo recordings utilizing a second hydrophone positioned along the sea wall 4 m to the left of the first one. Although two hydrophones are insufficient to localize a position, marked changes in level between the two channels would suggest that it is unlikely for fish to communicate over long distances.

The ceramic piezoelectric hydrophone and ac amplifier (designed and constructed by Robert F. Thorson) were battery powered and built with low-noise components. The

sounds were transmitted to a receiver via a Wavecom Sr, 2.4-GHz microwave transmission link, demodulated, and then stored digitally on the hard disk of a CTX FC3A300 computer and archived on CD. The hydrophone and amplifier were sensitive from 14 Hz (3-dB down point) to nearly 20 kHz. Most energy in toadfish calls is below 1 KHz,[9] and *Opsanus tau* hears to about 800 Hz.[17]

Boatwhistles were analyzed using a sampling rate of 11 020 Hz with SPECTRA-PLUS PROFESSIONAL version 4.0 software on the CTX FC3A300 computer. Because fish were of unknown distance from the hydrophone, sound-pressure levels were determined in dB relative to full scale on the monitor, and frequency spectra were determined using a 1024-point fast Fourier transform with a Hanning window.

As in previous work,[13] we identified boatwhistles of individual callers by a combination of their signal level, frequency spectra boop number, and duration. Grunts were also sufficiently stereotyped that we attempted to separate individuals similarly by a combination of frequency and amplitude on their frequency spectra (Figs. 1 and 2). For this analysis we obtained one representative frequency spectrum for each of the four fish weekly for 9 weeks. Since fish 4 did not call for 2 of these weeks, it was represented with seven spectra. Amplitude values were obtained from each spectrum at 50-Hz intervals and averaged for each fish. Data were analyzed with a two-way analysis of variance (ANOVA) with frequency and fish number as the two factors. Grunts were also separated into clusters by plotting the peak frequency against the amplitude at that frequency.

For analysis of tagging, each grunt was individually identified as coming from one of four nearby fish by determining its frequency spectrum and comparing it with representative templates from the individual fish. Latencies to tag, i.e., time from a sound (grunt or gruntlike portion of a boatwhistle) to grunt tag were measured in milliseconds using the oscilloscope display in the analysis software. In cases with multiple potential tag triggers (i.e., two separate grunts before a boatwhistle and the beginning gruntlike portion of the boatwhistle), we considered any tag occurring later than the shortest tag latency recorded from that fish to be triggered by that sound (see Sec. III for further clarification).

In order to demonstrate that tags are in fact triggered by an acoustic stimulus rather than being a random occurrence, we examined every grunt produced by fish 1 during the 2-h recording on 12 March 2000 and measured the time from the preceding stimulus that could have triggered the grunt. This interval could potentially vary from a small number of milliseconds to many seconds or even minutes. A histogram of this data (equivalent to a time-interval histogram in neurophysiology) will indicate whether grunts occur randomly in time or are tied to a specific stimulus.

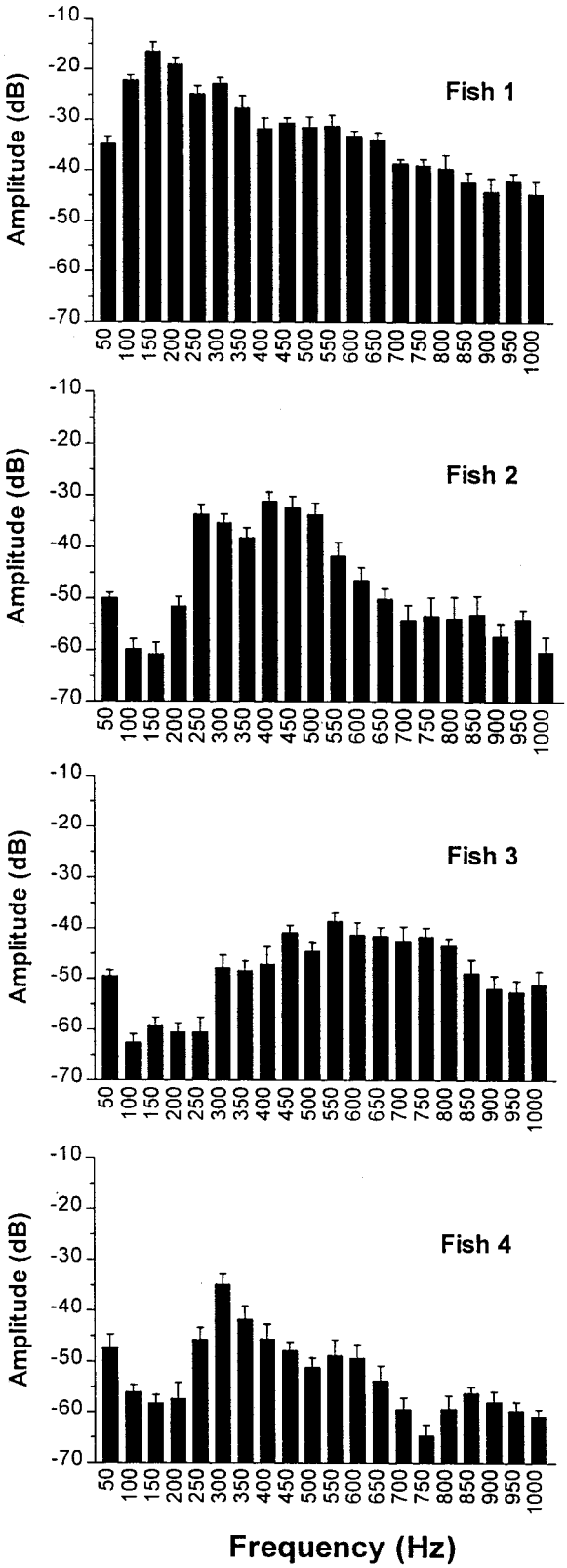


FIG. 1. Mean amplitude ± s.d. at 50-Hz intervals determined from frequency spectra of individual grunts obtained weekly over the course of 9 weeks for fish 1–4. $N=7$ for fish 4 who did not call during two of the recordings. Spectra were determined with a 1024-point fast Fourier transform and a Hanning window.

III. RESULTS

A. Individual identification

Four fish (fish 1–4) remained close enough to the hydrophone to be recognized repeatedly. Grunt frequency spectra are relatively stable for weeks, and differences in sound amplitude with frequency vary among fish and support separations based on boatwhistle parameters (Figs. 1 and 2). Comparison of the frequency spectra by fish number and frequency with two-way ANOVA is highly significant. There is a significant interaction accounting for approximately

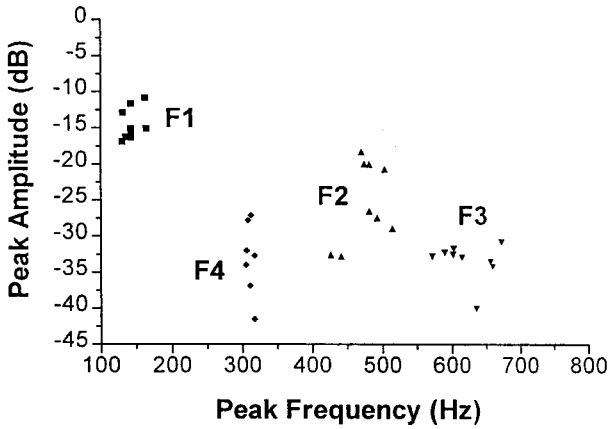


FIG. 2. Plot of peak amplitude in dB against the frequency of peak amplitude for individual grunt spectra of all four fish (separate symbols). Data points represent a single grunt per week recorded on 9 separate weeks. $N=7$ for fish 4 who did not call during two of the recordings.

22.85% of the variance ($F_{57,600}=10.62, p<0.0001$). Because of the interaction, row and column effects are difficult to interpret and will be given here as approximations. Frequency accounts for approximately 18.45% of the total variance ($F_{19,600}=25.72, p<0.0001$), and fish number accounts for 36.06% of the variance ($F_{3,600}=318.42, p<0.0001$), the largest of the three factors. For our current purposes it is unnecessary to compare the amplitude values for the four fish individually at each of the 20 frequencies, but it is clear that there are major differences in the frequency spectra among these individuals. Likewise, plots of the peak dB level against frequency separate the four fish into clusters with no overlap in peak frequency (Fig. 2). Although there is no independent confirmation of fish identity in field recordings, our separation of individuals is reasonable and parsimonious.

B. Tagging

We recognize tagging in about a third of the boatwhistles emitted during the crepuscular period during the two winter–spring seasons. Tagging is absent (2000) or rare (2001) during daylight hours. Note that the tag by fish 1 of fish 2 [Figs. 3(b) and (c)] has both lower frequency components (sonogram) and greater amplitude (oscillogram) than the grunt or the gruntlike beginning of the boatwhistle of fish 2. Fish also tagged snaps of snapping shrimp (Fig. 4), and both snaps and toadfish calls could set off chains of multiple tagging [Fig. 4(c)]. During the recording session, fish 1 tagged boatwhistles of all the other fish (132 tags), concentrated on fish 2 (128 tags), and shrimp (42 tags), but was tagged itself only twice. Fish 2 tagged 251 times and was tagged 149 times. Fish 3 and 4 tagged 20 and 3 times, respectively, and were tagged 234 and 18 times. The number of boatwhistles produced by fish 1–4 is relatively similar (235, 372, 282, and 197, respectively), but the number of untagged calls is higher in fish 1 and 2 (233 and 220) than in fish 3 and 4 (48 and 179). The percentage of tagged calls for the four is 0.9%, 41%, 83%, and 9%, respectively. Fish 4 produced only three tags, but it was tagged only 18 times. The smaller number of tagging interactions involving fish 4 may relate to its

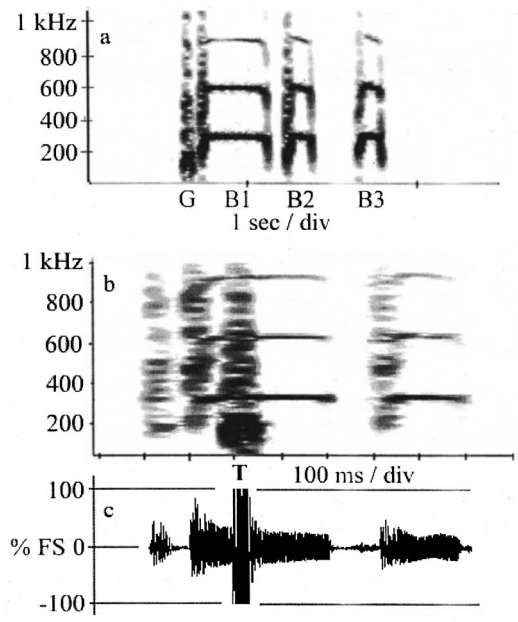


FIG. 3. Sounds of *Opsanus beta*. (a) Typical boatwhistle advertisement call with an initial grunt (G); a long tonal boop (B1) and two shorter boops (B2 and 3). (b) Sonogram and (c) Oscillogram of a boatwhistle produced by fish 2 that was tagged by fish 1. The T marks the tag, which has a greater amplitude and lower frequency energy than the call of fish 2.

location at a greater distance from the other fish, as suggested by the lower amplitude level evident in its frequency spectrum (Fig. 1). In early April 2001, toward the end of the mating season, three of the fish appeared to move from their original territories, causing their sound levels to change. Fish 1 ceased spontaneous calling on 24 April but continued to tag occasionally. At this time we encountered incidents of reciprocal tagging.

The decision to tag is made rapidly (Fig. 5). Latencies

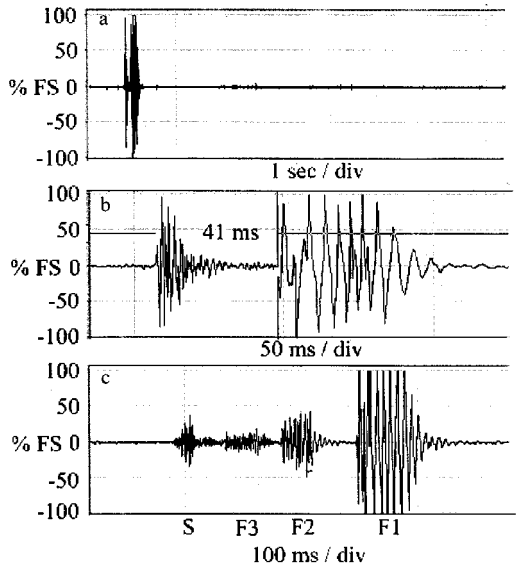


FIG. 4. Tags of shrimp snaps. (a) Oscillogram of a pistol shrimp snap tagged by fish 1 with a latency of 41 ms shown in real time. (b) Same selection expanded. (c) Chain of tags initiated by a shrimp snap that is tagged by fish 3. The fish 3 tag is then tagged by fish 2, who in turn is tagged by fish 1.

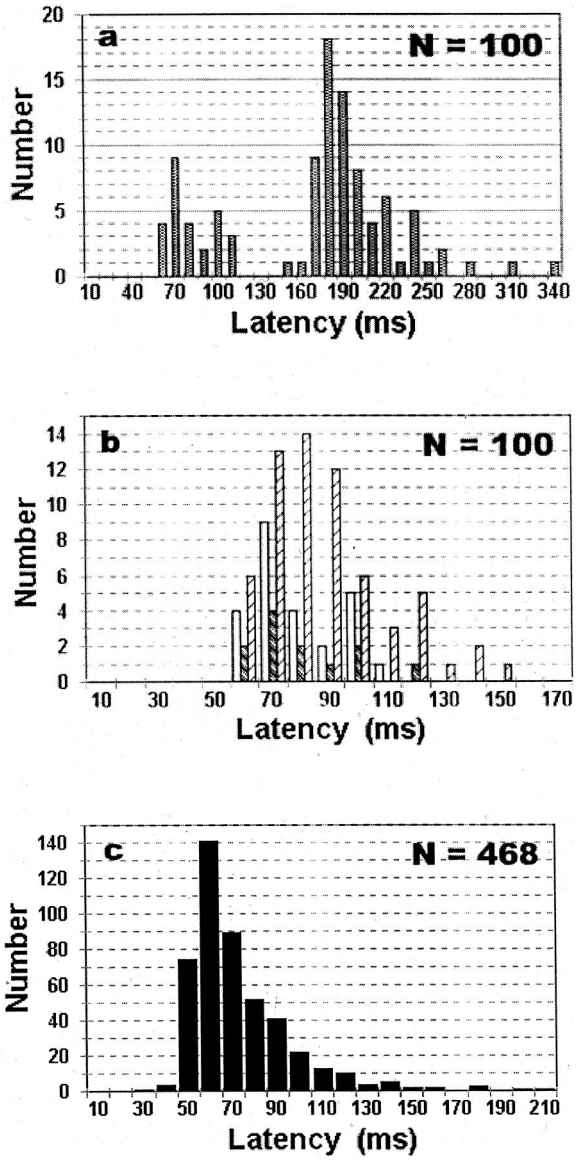


FIG. 5. Histogram of tag latencies on 12 March 2000. (a) Latency of fish 1 tags of fish 2 boatwhistles. (b) Latency of fish 1 tags after correction for triggering on grunt 1, grunt 2, or the initial gruntlike portion of the first boop. (c) Latency of all tags from the four fish after correcting for timing of trigger stimulus.

from the recording are as short as 41 ms for a pistol shrimp snap, which has a faster rise time than a toadfish grunt, and 53 ms for a grunt tagged by fish 1. Longest latencies are approximately a third of a second. The latency histogram of fish 1 tagging fish 2 is bimodal, with the first peak between 53–110 ms and a pause of over 43 ms before the start of the second mode [Fig. 5(a)]. Fish 2 sometimes produced two grunts before the gruntlike beginning of the boop. Using these three events as possible triggers, we estimate the latency by counting any tag occurring $\geqslant$53 ms (the latency of the shortest tagged grunt) as evoked by that trigger. With this correction almost all tags occur within the first time mode [Fig. 5(b)]. There are 25 tags after the first grunt, 12 after the second, and 63 after the gruntlike beginning of the boop. Similarly, earliest latencies to tag shrimp snaps and calls of other fish, respectively, are 43 and 44 ms for fish 2. 48 and

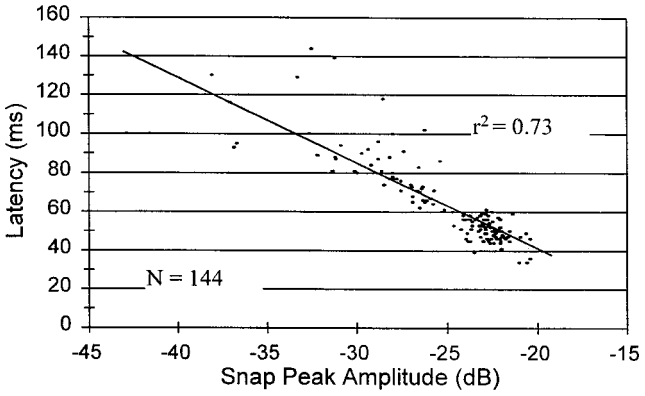


FIG. 6. Relationship of latency to peak amplitude for 144 shrimp snaps tagged by fish 1 on 21 March 2000. The regression equation is latency $= -46.13 - 4.36$ dB.

64 ms for fish 3, and 56 ms for fish 4 to tag a grunt (it did not tag shrimp snaps in this recording). After correcting for the trigger, latencies from all fish are similar, and average 68.7 $\pm$23.7 ms (s.d.) [Fig. 5(c)].

The trigger to tag is a relatively intense sound with a rapid rise time, a feature shared by toadfish grunts and shrimp snaps. Toadfish tend to tag more snaps when there are few boatwhistles (i.e., early in the crepuscular period before the toadfish increase their calling rate or on days with little calling). From a session with few calls (21 March 2000), we measured latencies of fish 1 to tag 144 shrimp snaps (Fig. 6). Snap amplitudes vary by over 20 dB, suggesting shrimp are in different positions relative to the hydrophone and therefore the focal fish. There is a linear relationship ($r^2 = 0.73$) between latency and amplitude so that latencies range from 34 to 144 ms for snaps ranging from -21 to -43 dB). Residuals from the regression line increase for weaker snaps, likely because the hydrophone was not immediately next to the fish.

Measurement of the intervals between each grunt produced by fish 1 and the preceding sound (i.e., potential trigger) in the 2-h recording indicate that tags occur immediately after a trigger stimulus and not randomly in time (Fig. 7). Of 178 grunts, 168 occur within 150 ms of a toadfish sound or a shrimp snap. Time intervals vary between 32 and 1630 ms, and the median interval is 73 ms. With 178 grunts in 2 h and

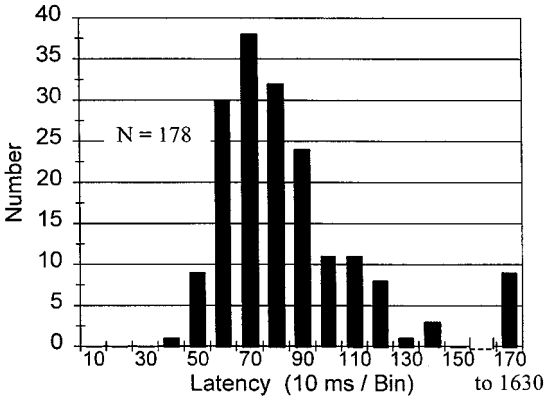


FIG. 7. Histogram of intervals between the preceding sound (grunt or snap) and all 178 grunts produced by fish 1 in 2-h recording on 12 March 2000.

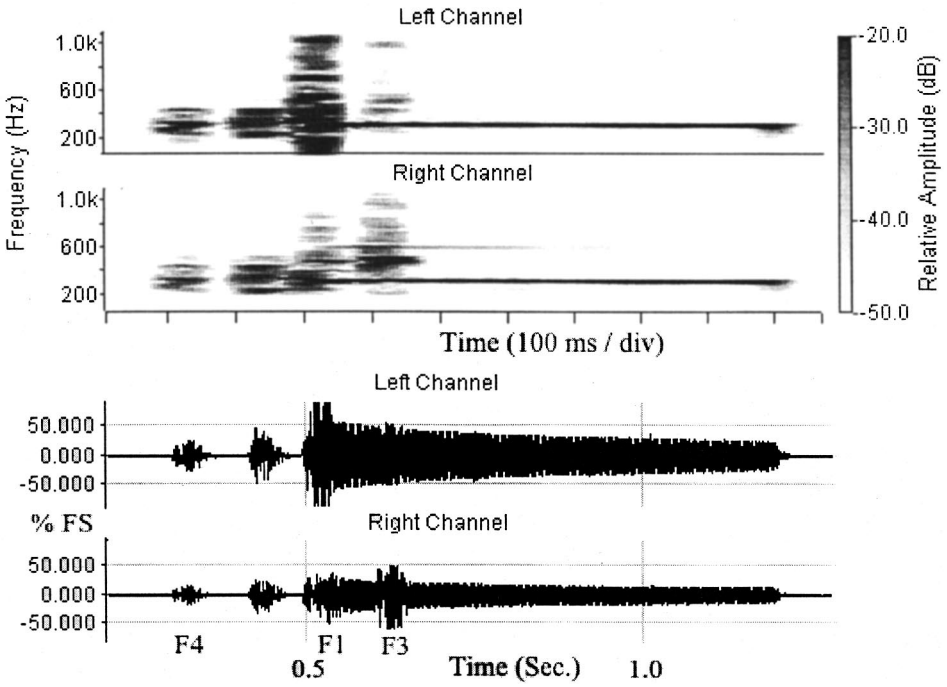


FIG. 8. Sonographs and oscillographs of two channel recordings (hydrophones 4 m apart) of two grunts and a long boop produced by fish 4 on 24 April 2001. The second grunt of fish 4 is tagged by fish 1 and is in turn tagged by fish 3. Note the disparity in sound amplitude, particularly for the tags produced by fish 1 and 3.

a time interval of 150 ms (a conservative value that includes 94% of the grunts by fish 1), the sum of all grunt intervals is 26.7 s out of 7200 s (120 min×60 s/min). Dividing the number of grunts into the session duration indicates that grunts would occur at an average of 40-s intervals if produced randomly. Clearly, these grunts (our tags) are triggered by an acoustic stimulus and are not a random occurrence.

Late season recording with two hydrophones (after three of the fish had moved from their original positions) indicates that fish 1, 2, and 4 were closer to hydrophone 1, and fish 3 was closer to hydrophone 2; representative differences in sound levels between the hydrophones for the four fish were 3, 4, 9, and 16 dB. Figure 8 illustrates a two-hydrophone recording of a chain of tags in which the second grunt of fish 4 is tagged by fish 1, who in turn is tagged by fish 3. The grunts of fish 4 are, respectively, 4.5 and 3.3 dB greater on hydrophone 1, and the tag by fish 1 is 9.6 dB greater on hydrophone 1. Fish 3, however, whose tag would not be recognized above the boatwhistle on the oscillogram from hydrophone 1, is at least 0.7 dB greater than the boatwhistle on hydrophone 2. A peak dB measurement of the highest amplitude of the most prominent frequency in the tag, which removes the background level of the boatwhistle, is 6.5 dB higher on hydrophone 2.

IV. DISCUSSION

The finding that gulf toadfish can grunt (tag) during the boatwhistle of another solves the riddle of boatwhistles recorded with embedded grunts in the oyster toadfish (*Opsanus tau*).[15] Latencies were not measured in the earlier study, but they appear to be longer in *O. tau* than *O. beta*, perhaps because the *O. tau* call is simpler, typically without preceding grunts that could cue the tagging fish. Gulf toadfish tag calls of nearby fish and snapping shrimp, and generally ignore less intense calls of toadfish located at a greater distance. The stimulus to tag appears to be an intense sound with a rapid rise time, and the decision to tag is usually made within 100 ms. The shrimp snap is an intense signal with a shorter rise time than the toadfish grunt [Fig. 4(b)], and the shortest latency was faster for a snap than for a toadfish grunt. Most of the energy in the shrimp snap is above the audible frequency range of toadfish (about 800 Hz in *Opsanus tau*),[17] but the snap's lower frequency components will be audible to the toadfish. We suggest that snaps are tagged because they mimic characteristics that stimulate the toadfish and not because of interspecific communication. Observations that toadfish tag shrimp snaps more frequently at times with few toadfish calls suggest an ability to discriminate snaps from grunts, a remarkable feat considering the short latency for a decision.

Since we cannot localize the position of the calling fish with two hydrophones, comparison of levels of sounds of the tagging and tagged fish are not clearly meaningful. However, a difference of as much as 16 dB between two hydrophones 4 m apart indicates rapid attenuation of these signals with distance in this shallow-water habitat. Further, a tag by fish 3 is almost completely masked in the oscillogram of the boop from hydrophone 1. Rapid attenuation of signals with distance in other toadfish[16,18] and a damselfish[19] likewise supports the notion that communication in *Opsanus beta* is restricted to short distances in shallow water.

Fish[10] apparently stimulated tagging experimentally in *Opsanus tau* without recognizing the phenomenon. Playbacks of boatwhistles or tone bursts at a rapid rate to calling toadfish will stimulate them to call faster,[11,12] but playbacks of continuous tones suppress calling.[10] Playbacks of two 6-s tone bursts with a silent interval between the bursts stimulate toadfish to produce a boatwhistle in the silent period. By shortening the interval between the bursts, Fish succeeded in getting toadfish to call with a ''minimum auditory time'' of 80 ms and more commonly response times of 120 to 140 ms. About 90% of the boatwhistles began in the silent interval

even when the intervals were only 0.26–0.32 s long. These playbacks also evoked grunts (i.e., our tags), which Fish[10] interpreted as aborted boatwhistles. Grunts became increasingly common as the intervals shortened, and 521 of 535 grunts occurred in the first 100 ms of the second tone burst, with most occurring within 70 ms, approximately the same time as our average latency to tag. We suggest that it is highly unlikely for a fish to hear a neighbor's boatwhistle clearly over his own call, react to that boatwhistle (or tone-burst playback), and then somehow abort a call (stop the motor message already descending from the forebrain)[20,21] in such a short time window, presuming of course that it can stop it at all.

The rapid reaction time for tagging and placing a boatwhistle in the silent interval between tones is supported by intimate connections of the auditory and motor pathways in toadfish.[22,23] Routine electrical stimulation of the mating call in the forebrain but not lower centers[20,21] suggests that the decision to tag is made at higher centers rather than being a reflex.

We believe that tagging is an undescribed phenomenon in animal communication. It is quite different from a facilitated response to other callers or playbacks since toadfish boatwhistles rarely overlap in the field.[10,13] It is also different from phenomena like production of the ''co'' note of the coqui frog,[24] which is directed at males but is part of the normal advertisement call because a male toadfish places his tag as a response to occur during the call of a rival male.

Since data come from field recordings of unseen fish, we have no independent observations that explain its function. However, the absence of reciprocal calling over most of the mating season suggests that the tagging fish is expressing dominance over the fish it tags; reciprocal tagging toward the end of the mating season suggests that reversals of status are possible. Tagging does not appear to decrease the number of boatwhistles produced by the tagged fish. A tag is also too short to mask or jam the boatwhistle of the tagged fish. Because a grunt requires several muscle contractions, and a boatwhistle can require several hundred,[13] tags are energetically efficient compared to a boatwhistle. Occurrence of tagging primarily during the twilight peak of calling[13,14] suggests the hypothesis that at least some of the information in the call is directed at females who would be using the boatwhistle advertisement call to choose a mate[12,25,26] during this peak period.

[1] M. L. Fine, K. L. Malloy, C. B. King, S. L. Mitchell, and T. M. Cameron, ''Movement and sound generation by the toadfish swimbladder,'' J. Comp. Physiol., A **187**, 371–379 (2001).

[2] L. C. Rome, D. A. Syme, S. Hollingworth, S. L. Lindstedt, and S. M. Baylor, ''The whistle and the rattle: The design of sound producing muscles,'' Proc. Natl. Acad. Sci. U.S.A. **93**, 8095–8100 (1996).

[3] C. R. Skoglund, ''Functional analysis of swimbladder muscles engaged in sound production of the toadfish,'' J. Biophys. Biochem. Cytol. **10**(Suppl), 187–200 (1961).

[4] M. P. Fish, ''Character and significance of sound production among fishes of the Western North Atlantic,'' Bull. Bingham Oceanogr. Coll. **14**, 1–109 (1954).

[5] G. A. Gray and H. E. Winn, ''Reproductive ecology and sound production of the toadfish *Opsanus tau*,'' Ecology **28**, 274–282 (1961).

[6] W. N. Tavolga, ''Underwater sounds produced by two species of toadfish *Opsanus tau* and *Opasnus beta*,'' Bull. Mar. Sci. **8**, 278–284 (1958).

[7] T. D. Waybright, U. Kollenkirchen, and M. L. Fine, ''Effect of size and sex on grunt production in the oyster toadfish,'' Soc. Neurosci. Abstr. **16**, 578 (1990).

[8] M. E. Dos Santos, T. Modesto, R. J. Matos, M. S. Grober, R. F. Oliveira, and A. Canario, ''Sound production by the Lusitanian toadfish, *Halobatrachus didactylus*,'' Bioacoustics **10**, 309–321 (2000).

[9] M. L. Fine, ''Seasonal and geographic variation of the mating call of the oyster toadfish *Opsanus tau*,'' Oecologia **36**, 45–57 (1978).

[10] J. F. Fish, ''The effect of sound playback on the toadfish,'' in *Behavior of Marine Animals, Vol. 2*, edited by H. E. Winn and B. Olla (Plenum, New York, 1972), pp. 386–434.

[11] H. E. Winn, ''Vocal facilitation and the biological significance of toadfish sounds,'' in *Marine Bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York, 1967), pp. 213–231.

[12] H. E. Winn, ''Acoustic discrimination by the toadfish with comments on signal systems,'' in *Behavior of Marine Animals: Current Perspectives in Research, Vol. 2. Vertebrates*, edited by H. E. Winn and B. L. Olla (Plenum, New York, 1972), pp. 361–385.

[13] R. F. Thorson and M. L. Fine, ''Crepuscular changes in emission rate and parameters of the boatwhistle advertisement call of the gulf toadfish, *Opsanus beta*,'' Environ. Biol. Fish **63**, 321–331 (2002).

[14] C. M. Breder, Jr., ''Seasonal and diurnal occurrences of fish sounds in a small Florida bay,'' Bull. Am. Mus. Nat. Hist. **138**, 325–378 (1968).

[15] M. L. Fine, ''Possible lateralization of function in toadfish sound production,'' Physiol. Behav. **25**, 167–169 (1982).

[16] M. L. Fine and M. L. Lenhardt, ''Shallow-water propagation of the toadfish mating call,'' Comp. Biochem. Physiol. **76A**, 225–231 (1983).

[17] H. Y. Yan, M. L. Fine, N. S. Horn, and W. E. Colon, ''Variability in the role of the gasbladder in fish audition,'' J. Comp. Physiol., A **187**, 371–379 (2000).

[18] A. H. Bass and C. Clark, ''The physical acoustics of underwater sound communication,'' in *Springer Handbook of Auditory Research*, edited by A. M. Simmons, A. N. Popper, and R. R. Fay (Springer, New York, 2002) (in press).

[19] D. A. Mann and P. S. Lobel, ''Propagation of damselfish (*Pomacentridae*) courtship sounds,'' J. Acoust. Soc. Am. **101**, 3783–3791 (1997).

[20] M. L. Fine and M. A. Perini, ''Sound production evoked by electrical stimulation of the forebrain in the oyster toadfish,'' J. Comp. Physiol., A **174**, 173–185 (1994).

[21] J. L. Goodson and A. H. Bass, ''Forebrain peptides modulate sexually polymorphic vocal circuitry,'' Nature (London) **403**, 769–772 (2000).

[22] A. H. Bass, D. A. Bodnar, and M. A. Marchaterre, ''Midbrain acoustic circuitry in a vocalizing fish,'' J. Comp. Neurol. **419**, 505–531 (2000).

[23] A. H. Bass, M. A. Marchaterre, and R. Baker, ''Vocal-acoustic pathways in a teleost fish,'' J. Neurosci. **14**, 4025–4039 (1994).

[24] P. M. Narins and R. R. Capranica, ''Communicative significance of the two-note call of the treefrog, *Elutherodactylus coqui*,'' J. Comp. Physiol. **127A**, 1–9 (1978).

[25] R. K. Brantley and A. H. Bass, ''Alternative male spawning tactics and acoustic signals in the plainfin midshipman fish *Porichthys notatus* Girard (*Teleostei, Batrachoididae*),'' Ethology **96**, 213–232 (1994).

[26] J. R. McKibben and A. H. Bass, ''Behavioral assessment of acoustic parameters relevant to signal recognition and preference in a vocal fish,'' J. Acoust. Soc. Am. **104**, 3520–3533 (1998).

Frequency-dependent attenuation-compensation functions for ultrasonic signals backscattered from random media

Michael L. Oelze[a] and William D. O'Brien, Jr.
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

(Received 18 July 2001; accepted for publication 24 December 2001)

Estimations of scattering parameters, such as average scatterer diameter, from rf signals backscattered from random media (tissues) are made from the frequency dependence of the rf signal. The frequency dependence of the rf signal backscattered from the medium is seen in the normalized power spectrum. The normalized power spectrum is found by taking the squared magnitude of the Fourier transform of the rf signal gated over a region of interest and dividing by some reference spectrum. If the medium has a frequency-dependent attenuation then the shape of the normalized power spectrum will be affected by the frequency-dependent attenuation and the time duration of the gated signal. Not accounting for the frequency-dependent attenuation leads to poor estimations of scatterer parameters. Larger attenuation and longer time gates give poorer estimates of scatterer parameters without attenuation compensation. Several attenuation-compensation functions have been used to account for the attenuation losses to the normalized power spectrum. A new attenuation-compensation function is proposed and compared with the other attenuation-compensation routines. The new attenuation-compensation function is shown to give improved estimates over previous attenuation-compensation functions for scatterers that follow a Gaussian form factor. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1452743]

PACS numbers: 43.80.Vj, 43.80.Qf [FD]

I. INTRODUCTION

Clinical ultrasound B-mode images of tissues aid in diagnosis of disease. B-mode images are constructed from envelop-detected backscattered echoes that may contain information about the structure of the biological media. The frequency spectra of rf backscattered echos have been used to noninvasively investigate and parametrize small structures in tissues.[1–9] Information about the structures can be obtained by relating the frequency dependence of the measured spectrum to theoretical predictions.

In many cases the medium (tissues) can be approximated as a random distribution of small scatterers or particles. The frequency spectrum of signals scattered from these particles depends on their average size, shape, distribution, and acoustic impedance. Models have been developed to predict the interaction of sound with a random distribution of scatterers.[10,11] Commonly, tissue scattering theory is described by spatial correlation function models.

The spatial correlation function models assume weak scattering theory, i.e., the incident wave is scattered only once and multiple scattering from particles is assumed to be negligible. A theoretical scattered power spectrum is related to the spatial correlation function through an acoustic intensity form factor.[7] The form factor is determined by evaluating the Fourier transform of the correlation function for the tissue medium. The form (or shape) factor gives the frequency dependence of the normalized power spectrum from the scattering in terms of the assumed shape and distribution of the scatterers.

The frequency dependence of the normalized power spectrum measured from the backscattered echoes can be related to the form factor. The average effective scatterer size and concentration of scatterers can be estimated from the relation between the normalized power spectrum and the modeled form factor. If the attenuation is not negligible and if the attenuation has a frequency dependence, which is almost always the case, then a compensation function is needed to account for the frequency-dependent losses to the normalized power spectrum. As longer gated segments are used to obtain the normalized power spectrum, the effects of frequency-dependent attenuation increases. Assuming that scattering is homogeneous over the length of the gated segment (stationary), then on average the difference between wavelets scattered near the beginning of the time gate and those scattered near the end will be the frequency-dependent losses. If the effects of frequency-dependent attenuation are not properly taken into account, estimations of scatterer properties from the normalized power spectrum will be incorrect.

This work examines the effects of using different attenuation-compensation functions to account for losses to the normalized power spectrum taken from gated signals. Frequency-dependent attenuation losses come from propagation between the gated region and the source (intervening medium) and over the length of the gated region itself. The accuracy of scatterer estimates depends on the compensation of both propagation regions. This paper focuses on the compensation of attenuation losses to backscattered signals over

[a]Electronic mail: oelze@brl.uiuc.edu

the gated length. A new attenuation-compensation function is proposed and compared with existing attenuation-compensation functions to determine the best compensation function to use in a particular attenuating media. Section II outlines the theory describing the development of a new attenuation-compensation function. Section III compares the new attenuation-compensation function with other attenuation-compensation functions. Section IV examines the attenuation-compensation functions for large attenuation or long gate length. Section V gives some concluding remarks about the compensation of backscattered signals.

II. ATTENUATION-COMPENSATION TECHNIQUES

A. Derivation of attenuation-compensation functions

Scatterer properties of random scattering media are estimated from the frequency dependence of the normalized power spectrum through the form factor. One form factor that has been shown to model the scattering from many soft tissues well is the Gaussian form factor[2,4–8,12]

$$F_{\text{Gauss}}(2k) = e^{-0.827k^2 a_{\text{eff}}^2}, \tag{1}$$

where k is the acoustic wave number, and a_{eff} is the average effective radius of the scatterers. The Gaussian form factor describes spherical scattering sources with impedance that varies continuously from the surrounding tissue. The average scatterer size can be estimated from measurements of the form factor.

The form factor is defined as the ratio of the backscatter coefficient, σ_b, for the scattering medium to the backscatter coefficient of a medium consisting of point scatterers,[6]

$$F(2k) = \frac{\sigma_b}{\sigma_0}, \tag{2}$$

where σ_0 is backscatter coefficient for point scatterers. Equation (2) gives a theoretical description of the scattering that can be related to actual acoustic measurements through the normalized power spectrum. The normalized power spectrum is related to the form factor through the backscatter coefficient[6]

$$\sigma_b = \frac{CR^2}{A_0 L} W_{\text{meas}}(f), \tag{3}$$

where C is a constant used to normalize the effects of a gating function on the backscattered echo and the effects of the beam of the transducer, A_0 is the area of the transducer face, L is the length of the gated segment, R is the on-axis distance between the transducer and the proximal surface of the gated region, and $W_{\text{meas}}(f)$ is the normalized power spectrum. Typically, the normalized power spectrum is measured from a region of interest (ROI) and is an average of the spectral contributions from rf time signals within the specified region. The measured power spectrum from a single rf line in the ROI is the magnitude squared of the Fourier transform, FT(f), of the backscattered time signal gated according to the length, L, of the ROI. The normalized power spectrum is given by averaging each measured power spectrum in the ROI and dividing by a reference spectrum[7,13]

$$W_{\text{meas}}(f) = \frac{A(f,L)}{N} \sum_{m=1}^{N} \frac{|\text{FT}_m(f)|^2}{|\text{FT}_{\text{ref}}(f)|^2} = A(f,L) W_{\text{atten}}(f), \tag{4}$$

where N is the number of gated segments to be averaged and A is an attenuation-compensation function. The reference spectrum takes out the effects of the equipment on the measurement and is found by the substitution technique.[1,14] The attenuation-compensation function is unitless and depends on the frequency-dependent attenuation coefficient (Np/cm) in the gated length, the length of the gated signal, and the propagation losses in the intervening medium. The frequency dependence of the normalized power spectrum is used to make estimates of average scatterer size.

Accounting for the effects of frequency-dependent attenuation is necessary to accurately estimate scatterer properties such as the average scatterer diameter and concentration. Point attenuation compensation has been used in the past to compensate for signal loss when the gated segment is short or the attenuation is small.[6,7] The normalized power spectrum is compensated for round-trip attenuation losses using point compensation by

$$W_{\text{meas}}(f) = W_{\text{atten}}(f) e^{4\alpha_0(f)x_0} e^{4\alpha(f)L/2}, \tag{5}$$

where x_0 is the distance between the source and the gated region, $L/2$ is the distance from the edge of the gated region to the middle of the gated region, and $\alpha_0(f)$, $\alpha(f)$ are the frequency-dependent attenuation coefficients in Np/cm for the intervening medium and over the gated region, respectively. The attenuation over the gated region is assumed to be constant.

The simulations in Sec. III will show that when longer gated segments are used or the medium has higher attenuation, i.e., $\alpha(f)L > 0.5$, point compensation does not correctly account for the frequency-dependent effects of attenuation on the normalized power spectrum. The justification for using longer gate lengths is that better estimates of scatterer properties can be obtained.[15] Chen *et al.* compared experimental measurements of backscatter coefficients (power spectrum) from phantoms of glass beads in agar using gate lengths of 3 mm, 7.7 mm, and 3 cm with theoretical predictions.[15] The results of the experiments by Chen *et al.* showed that the longer gate lengths followed more closely the theoretical spectrum over the largest frequency range. When $\alpha(f)L > 0.5$, point attenuation compensation tends to overcompensate for the frequency dependence of the normalized power spectrum. Overcompensating for the frequency-dependent losses causes the underestimation of average scatterer diameters. Figure 1 shows how the overcompensation of frequency-dependent attenuation leads to underestimation of scatterer sizes. From Fig. 1 it is seen that the form factor for smaller scatterers has a flatter slope than for larger scatterers. Typically, attenuation increases with increasing frequency. If an attenuation-compensation function overcompensates, then the large frequencies will be overcompensated more than the lower frequencies. The net effect is to raise the overall slope of the measured form factor causing an underestimation of the scatterer size.

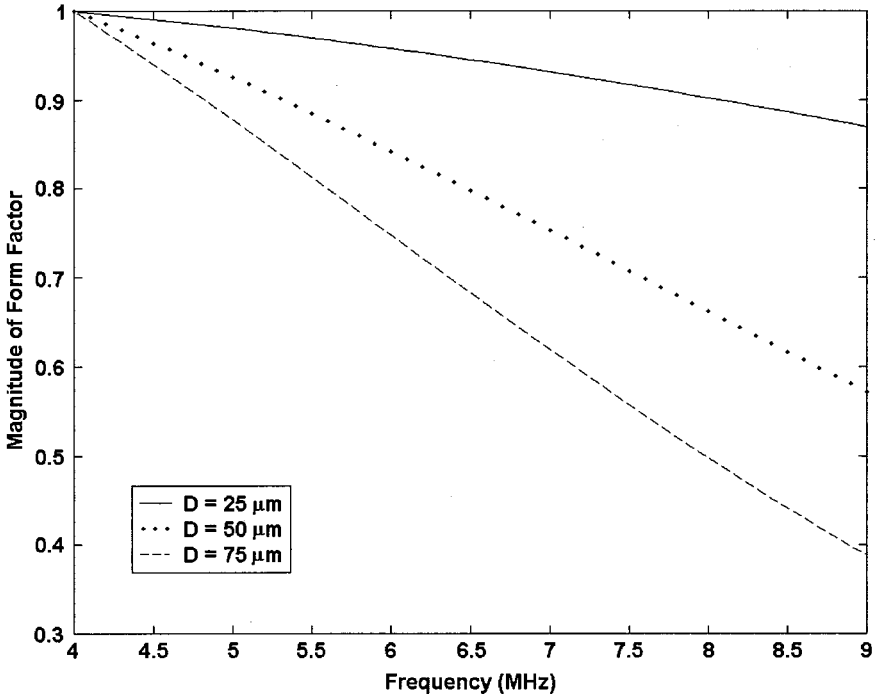


FIG. 1. The relative magnitude of the Gaussian form factor over the frequency range of 4–9 MHz for scatterers with diameters of 25, 50, and 75 μm.

In order to obtain more accurate estimates of scatterer diameters, the effects of larger attenuation or attenuation from longer gated signals need to be more accurately compensated. An approximate closed form attenuation-compensation function, A, for backscattered returns was first introduced by Sigelmann and Reid[14] and utilized by O'Donnell and Miller.[16] The O'Donnell and Miller attenuation-compensation term is given by

$$W_{\text{meas}}(f)=W_{\text{atten}}(f)A_{\text{OM}}(f,L),$$

where

$$A_{\text{OM}}(f,L)=e^{4\alpha_0(f)x_0}\left[\frac{4\alpha(f)L}{1-e^{-4\alpha(f)L}}\right]. \tag{6}$$

The term in square brackets accounts for the frequency-dependent attenuation losses over the gated region. The term outside the square brackets accounts for the round-trip frequency-dependent attenuation losses between the gated region and the source. The O'Donnell and Miller compensation function is based on the approximation of small attenuation or short gate length over the frequencies of interest.[16] The simulations of Sec. III will show that the compensation term used by O'Donnell and Miller gives better estimations for scatterer estimations than the point attenuation-compensation term at longer gate lengths and larger attenuations. However, as $\alpha(f)L$ becomes larger the O'Donnell and Miller compensation function begins to yield increasingly inaccurate estimates of scatterer diameters.

To improve scatterer estimations, an alternative attenuation-compensation function is proposed. Consider a signal of the form, $g_L(t)$, representing a rf time sequence from a gated length, L, in a scattering medium. The function, $g_L(t)$, is the convolution of an impulse response, $p(t)$ (incorporating the electromechanical characteristics of the transducer and diffraction), and a scattering function, $r_L(t)$, from randomly spaced, identical particles,[17,18]

$$g_L(t)=p(t)*r_L(t). \tag{7}$$

The Fourier transform of the signal is given by

$$G_L(f)=P(f)R_L(f). \tag{8}$$

The Fourier spectrum, $R_L(f)$, over the rectangular gated length, L, is defined as

$$R_L(f)=\int_0^{2L/c} r_L(t)e^{-i2\pi ft}\,dt \tag{9}$$

and can be broken into N discrete integrals of length dx,

$$\begin{aligned}R_L(f)=&\int_0^{2\,dx/c} r_L(t)e^{-i2\pi ft}\,dt\\&+\int_{2\,dx/c}^{2(2)dx/c} r_L(t)e^{-i2\pi ft}dt+\cdots\\&+\int_{2(N-1)dx/c}^{2N\,dx/c} r_L(t)e^{-i2\pi ft}\,dt.\end{aligned}$$

Letting $r_{dx}(t)$ represent the scattering function of randomly spaced identical scatterers from some small dx segment then

$$\begin{aligned}r_{dx}(t)=&s(t+t_0)+s(t+t_0+2d_2/c)\\&+s(t+t_0+2d_3/c)+\cdots,\end{aligned} \tag{10}$$

where c is the speed of sound in the medium and d_i represents the individual scatterer spacings. The Fourier transform of $r_{dx}(t)$ is then given by

$$\begin{aligned}R_{dx}(f)=&S(f)\\&\times[1+e^{i2\pi f(2d_1/c)}+e^{i2\pi f(2d_2/c)}+\cdots]e^{-i2\pi t_0},\\R_{dx}(f)=&S(f)[1+\Phi(f,x)]e^{i2\pi ft_0},\end{aligned} \tag{11}$$

where $\Phi(f,x)$ is a random phase caused by the interference between the scattering from particles in the dx section located at a position, x. Essentially, $\Phi(f,x)$ sums together all of the random phase contributions of the identical particles caused by their random spacings. For a small number of particles it is expected that $\Phi(f,x)\neq 0$ but as the number of particles, N_p, becomes large, $\Phi(f,x)\rightarrow 0$. For a particular number density of scatterers within a certain volume then as the volume increases, N_p also increases and $\Phi(f,x)\rightarrow 0$.

Assuming that consecutive dx sections are small and are homogeneous with scattering (similar number density of scatterers and scatterer type) then the round-trip propagation loss can be incorporated in Eq. (11) by

$$R_{dx}(f)\approx S(f)[1+\Phi(f,x)]e^{i2\pi ft_0}e^{-2\alpha(f)x}, \tag{12}$$

where $\alpha(f)$ is the frequency-dependent attenuation in the gated region (assumed constant). The spatial average Fourier spectrum of the scattering function from the length, L, is the sum of all the dx segments in the total length divided by L,

$$R_L(f)=S(f)e^{-i2\pi ft_0}\frac{1}{L}\int_0^L[1+\Phi(f,x)]e^{-2\alpha(f)x}dx. \tag{13}$$

The second integral term involving $\Phi(f,x)$ is a summation of the random phase points of the identical scatterers in the gated segment, L, weighted by the attenuation of the signal at the depth, x. The attenuation diminishes the contribution of scatterers from greater propagation distances in the gated length. The attenuation effectively reduces the number of scatterers contributing to the phase summation. If the effective number of scatterers in the segment, L, is large then the integral term involving $\Phi(f,x)$ will be negligible. Integration of the first term yields

$$R_L(f)=S(f)e^{i2\pi ft_0}\frac{1-e^{-2\alpha(f)L}}{2\alpha(f)L}. \tag{14}$$

The power spectrum for the gated segment of length, L, is defined as

$$W_{\text{meas}}(f)=|G_L(f)|^2e^{-4\alpha_0(f)x_0}, \tag{15}$$

where the exponential term accounts for the round-trip propagation loss of the signal in the intervening medium. Substituting Eqs. (8) and (14) into Eq. (15) gives

$$W_{\text{meas}}(f)=|P(f)|^2|S(f)|^2\frac{(1-e^{-2\alpha(f)L})^2}{4\alpha^2(f)L^2}e^{-4\alpha_0(f)x_0}=\frac{|P(f)|^2|S(f)|^2}{A_{\text{OO}}(\alpha,L)}, \tag{16}$$

where $A_{\text{OO}}(\alpha,L)$ is an attenuation-compensation function (Oelze and O'Brien compensation) for signals gated from a scattering medium.

The O'Donnell and Miller compensation function can be derived by further examination of Eq. (11). The power spectrum of $R_{dx}(f)$ for some small section, dx, is given by

$$|R_{dx}(f)|^2=|S(f)|^2|[1+\Phi(f,x)]|^2. \tag{17}$$

Incorporating the round-trip propagation loss yields

$$|R_{dx}(f)|^2\approx|S(f)|^2|[1+\Phi(f,x)]|^2e^{-4\alpha(f)x} \tag{18}$$

for a particular dx segment. The spatial average of the scattered power spectrum from some length, L, is the average of the spectra of each dx segment giving

$$|R_L(f)|^2=\frac{|S(f)|^2}{L}\int_0^L|[1+\Phi(f,x)]|^2e^{-4\alpha(f)x}dx. \tag{19}$$

Expanding the terms yields

$$|R_L(f)|^2=\frac{|S(f)|^2}{L}\int_0^L(1+[\Phi+\Phi^*]+\Phi^*\Phi)e^{-4\alpha(f)x}dx. \tag{20}$$

The second and third terms can be considered negligible if a large effective number of scatterers exist in L. The normalized power spectrum from the gated length, L, will then be given by Eq. (15),

$$W_{\text{meas}}(f)=|P(f)|^2|S(f)|^2\frac{1-e^{-4\alpha(f)L}}{4\alpha(f)L}e^{-4\alpha_0(f)x_0}=\frac{|P(f)|^2|S(f)|^2}{A_{\text{OM}}(\alpha,L)}. \tag{21}$$

Two limiting cases of the attenuation-compensation term are when the quantity $\alpha(f)L$ is small. In the case of small $\alpha(f)L$, keeping only first-order terms in the exponential terms of A_{OO} and A_{OM} gives

$$1-e^{-2\alpha(f)L}\approx 2\alpha(f)L$$

and

$$1-e^{-4\alpha(f)L}\approx 4\alpha(f)L,$$

which reduces A_{OO} and A_{OM} to

$$A_{\text{OO}}(f,L)=A_{\text{OM}}(f,L)\approx e^{4\alpha_0(f)x_0} \tag{22}$$

or the point attenuation-compensation term. Keeping only the second-order terms yields

$$A_{\text{OO}}(f,L)=A_{\text{OM}}(f,L)\approx e^{4\alpha_0(f)x_0}\left[\frac{1}{1-2\alpha(f)L}\right]. \tag{23}$$

To second order the Oelze and O'Brien attenuation-compensation function and the O'Donnell and Miller function agree.

B. Error analysis

The difference between the Oelze and O'Brien attenuation-compensation function and the O'Donnell and Miller attenuation-compensation function can be seen by examination of the random phase summation terms that were approximated to be negligible. Neglecting the intervening medium compensation term and looking at the average power spectrum from Eq. (15) including the phase summation terms shows

$$W_{\text{OO}}(f)=\frac{|P(f)|^2|S(f)|^2}{L^2}\times\left|\int_0^L[1+\Phi(f,x)]e^{-2\alpha(f)x}dx\right|^2, \tag{24}$$

$$W_{\mathrm{OO}}(f)=\frac{|P(f)|^2|S(f)|^2}{L^2}\times\left(\frac{1}{A_{\mathrm{OO}}}+\frac{L}{A_{\mathrm{OO}}^{1/2}}\int_0^L(\Phi+\Phi^*)e^{-2\alpha(f)x}dx+\left|\int_0^L\Phi e^{-2\alpha(f)x}dx\right|^2\right)$$

The "error" terms of the Oelze and O'Brien attenuation-compensation function are given by the second and third terms of Eq. (24),

$$\epsilon_{\mathrm{OO}}(f)=\frac{|P(f)|^2|S(f)|^2}{L}\left(\frac{1}{A_{\mathrm{OO}}^{1/2}}\int_0^L(\Phi+\Phi^*)e^{-2\alpha(f)x}dx+\frac{1}{L}\left|\int_0^L\Phi e^{-2\alpha(f)x}dx\right|^2\right). \tag{25}$$

If α is large then $A_{\mathrm{OO}}\rightarrow 4\alpha^2L^2$, giving

$$\epsilon_{\mathrm{OO}}(f)=\frac{|P(f)|^2|S(f)|^2}{L}\left(\frac{1}{2\alpha L}\int_0^L(\Phi+\Phi^*)e^{-2\alpha(f)x}dx+\frac{1}{L}\left|\int_0^L\Phi e^{-2\alpha(f)x}dx\right|^2\right). \tag{26}$$

For the O'Donnell and Miller attenuation-compensation function the error terms can be seen from Eq. (20),

$$\epsilon_{\mathrm{OM}}(f)=\frac{|P(f)|^2|S(f)|^2}{L}\left(\int_0^L(\Phi+\Phi^*)e^{-4\alpha(f)x}dx+\int_0^L|\Phi|^2e^{-4\alpha(f)x}dx\right). \tag{27}$$

One would expect $\Phi(f,x)\rightarrow 0$ for a large number of scatterers. For a small number of random scatterers $\Phi(f,x)$ could not be assumed zero. The larger the gated segment the more likely $\Phi(f,x)\rightarrow 0$.

The trade-off between smaller gate size and better scatterer resolution can be seen from the error terms. The estimates of scatterer sizes are obtained from the term $|S(f)|^2$. Knowing the impulse response, $|P(f)|^2$, and the attenuation-compensation functions, the approximate term, $|S(f)|^2$, can be found from Eq. (16) or (21) by assuming that terms involving $\Phi(f,x)$ are small. However, if a small gate length means few random scatterers, the effects of attenuation may be negligible but the effects of the random scattering spacings on the measured power spectrum will be larger because $\Phi(f,x)\neq 0$. If a large gate length means many random scatterers, the greater the attenuation effect but the greater the likelihood that $\Phi(f,x)\rightarrow 0$. On the other hand, if the attenuation is large such that the effective number of scatterers is reduced to a few, the longer gate length will be of no advantage. What attenuation does is to weight the contribution of the phase points from scatterers at greater depth less than scatterers located at a shallower propagation depth. Attenuation effectively reduces the number of scatterer particles contributing to $\Phi(f,x)$. The effective reduction in scatterer number means it is less likely that $\Phi(f,x)$ will have negligible contribution to the power spectrum from a signal gated from length, L. Hence, estimation of scatterer sizes will be more accurate when the attenuation is small as opposed to large.

Comparing the error terms for the Oelze and O'Brien development [Eq. (26)] and the O'Donnell and Miller compensation function [Eq. (27)] shows that for large attenuation the first error term of the Oelze and O'Brien compensation is less than the first error term of the O'Donnell and Miller compensation function. The first terms are identical except for the $1/[2\alpha L]$ factor in the Oelze and O'Brien error term and the factor of 1/2 in the power of the exponent. The $1/[2\alpha L]$ factor and the reduced power in the attenuation exponent of the Oelze and O'Brien error term means that for large attenuation, $\alpha(f)L>1$, the first term of the Oelze and O'Brien term is much less than the O'Donnell and Miller error term. Comparisons of the second error terms also show that the Oelze and O'Brien error is less. Integrating $|\Phi|^2$ over the gate length means that all the summation of the phase points must be greater than or equal to zero. Thus, even without the attenuation and with a large number of scatterers the second term of the error term, Eq. (27), will not sum to zero. Comparisons of the error terms show that the Oelze and O'Brien attenuation-compensation function is superior to the O'Donnell and Miller attenuation-compensation function.

Since the second integral term in Eq. (27) is always positive, the O'Donnell and Miller term will undercompensate the attenuated, gated signal. For a larger gate length, i.e., more scatterers, the second error term of the O'Donnell and Miller attenuation-compensation function increasingly undercompensates the attenuated, gated signal. If the attenuation is larger, then the effective number of scatterers contributing to the undercompensation is less. Thus, the contribution of the second error term in the O'Donnell and Miller attenuation-compensation function is lessened with larger attenuation but increased with longer gate length. In either case, the second term means the O'Donnell and Miller attenuation-compensation function will undercompensate the attenuation losses in a gated signal from random scatterers.

Figure 2 shows the magnitude of three different compensation functions with increasing $\alpha(f)L$. Increasing $\alpha(f)L$ may result from increasing the rectangular gate length or from increasing the attenuation with frequency. Figure 2 shows that the point attenuation-compensation function gives the greatest overall compensation. The O'Donnell and Miller compensation term gives the smallest overall compensation to the attenuation losses. The Oelze and O'Brien compensation term gives an adjustment to the attenuation losses that lies between the point compensation and the O'Donnell and Miller compensation term. As seen in Fig. 2, both the O'Donnell and Miller compensation term and the Oelze and O'Brien compensation term reduce to the point compensation term for small $\alpha(f)L$.

C. Effects of windowing

The derivations of the attenuation-compensation functions assumed a rectangular window function was applied to the gated time signal. In many applications commercial win-

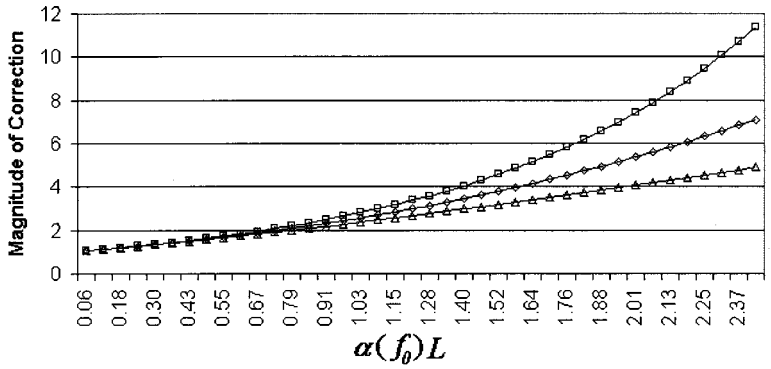


FIG. 2. Comparison of the magnitude of the correction to the normalized power spectrum using three different attenuation-compensation routines; (□) point compensation; (△) O'Donnell and Miller compensation; (◇) Oelze and O'Brien compensation.

dow functions other than the rectangular window are used to gate rf time signals. A window function applied to the time sequence, $g_L(t)$, can be represented by

$$g_L'(t)=h_L(t)g_L(t)=h_L(t)[p(t)*r_L(t)], \tag{28}$$

where $h_L(t)$ is the windowing function. The Fourier transform of the sequence $g_L'(t)$ is given as

$$G_L'(f)=H(f)*[P(f)R(f)], \tag{29}$$

where the Fourier spectrum of the window function is convolved with the impulse spectrum and the scattered signal spectrum. The convolution with the scattered spectrum means that the window function spectrum will also be convolved with the attenuation-compensation function derived using a rectangular window.

To obtain the attenuation-compensation function for any arbitrary window function it is necessary to deconvolve the window spectrum from the attenuation-compensation function derived using a rectangular window. The deconvolution is carried out by examining some small segment, *dx*, of the windowed signal, $g_L'(t)$. For a small *dx* segment the window function varies slowly compared to the scattered signal. The window function can then be approximated as a rectangular window with amplitude, $h(x)$, dependent on the position, *x*, of the *dx* element. Following Eq. (12), the Fourier transform of the *dx* segment with the attenuation is

$$G_{dx}'(f)\approx P(f)S(f)h(x)[1+\Phi(f,x)]e^{-2\alpha(f)x}. \tag{30}$$

The average, windowed power spectrum is obtained by integrating the *dx* segments over the gate length, *L*, giving

$$\langle G_{dx}'(f)\rangle\approx\frac{P(f)S(f)}{L}\times\int_0^L h(x)[1+\Phi(f,x)]e^{-2\alpha(f)x}dx. \tag{31}$$

Assuming that the second integral term is negligible then the attenuation function adjusted for the window shape is found by integration of the first term.

A popular window function that is often used is the Hanning window defined as

$$h(x)=\begin{cases}\dfrac{N}{2}\left[1-\cos\left(\dfrac{2\pi x}{L}\right)\right], & 0\leqslant x\leqslant L\\ 0 \quad \text{otherwise}\end{cases}, \tag{32}$$

where *N* is a normalization factor. Insertion of the Hanning window into Eq. (31) and integration of the first integral term yields

$$\int_0^L h(x)e^{-2\alpha(f)x}dx=\frac{N}{2}\left[\frac{1-e^{-2\alpha(f)x}}{2\alpha(f)L\left(1+\left[\dfrac{2\alpha(f)L}{2\pi}\right]^2\right)}\right]. \tag{33}$$

The attenuation-compensation function for the gated signal using a Hanning window follows from Eq. (16) giving

$$A_{\mathrm{OO}_{\mathrm{Han}}}(f,L)=A_{\mathrm{OO}}(f,L)\cdot\left[\frac{2}{N}\left(1+\left[\frac{2\alpha(f)L}{2\pi}\right]^2\right)\right]^2. \tag{34}$$

The normalization factor is found by noting that the attenuation-compensation function reduces to $e^{4\alpha_0(f)x_0}$ as the gate length goes to zero. As $L\to 0$, $A_{\mathrm{OO}}\to e^{4\alpha_0(f)x_0}$ and $A_{\mathrm{OO}_{\mathrm{Han}}}$ becomes

$$A_{\mathrm{OO}_{\mathrm{Han}}}(f,L)=\left(\frac{2}{N}\right)^2 e^{4\alpha_0(f)x_0}. \tag{35}$$

Setting Eq. (35) equal to $e^{4\alpha_0(f)x_0}$ gives the normalization factor for the Hanning window

$$N=2.$$

Simulations of scattered signals from Gaussian-type scatterers were also run using the Hanning window. Estimates of scatterer size were made using the Oelze and O'Brien attenuation-compensation function with the Hanning window adjustment and compared with estimates using other attenuation-compensation algorithms.

III. SIMULATION RESULTS

Numerical simulations were constructed to test the ability of the attenuation-compensation algorithms to account for the frequency-dependent losses and to determine the best algorithm to use. The numerical simulations were made by constructing a random matrix of scattering particles within a length of propagating medium of speed 1540 m/s. A transmitted pulse was constructed as the reference pulse, shown in Fig. 3. The reference pulse had a central frequency of about 8 MHz.

From each point in the matrix where a scattering particle was located, the reference pulse was scattered by multiplying the reference pulse by the form (or shape) factor of the scatterer. The form factor used in the simulations was the Gauss-

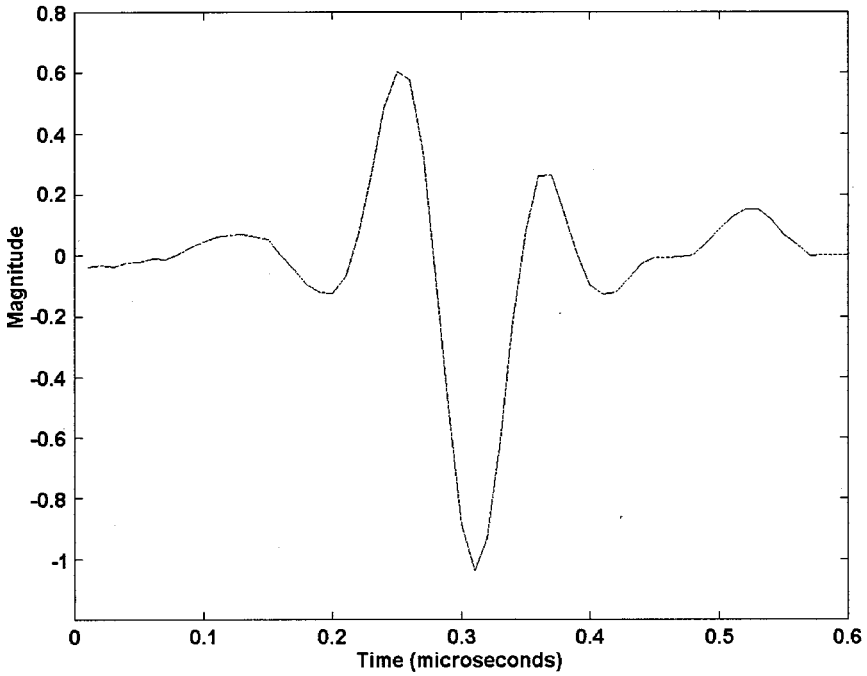


FIG. 3. Reference pulse used in the random scatterer simulations.

ian form factor. The simulated particles scattered according to the Gaussian form factor with an average effective diameter of 49 μm. An average diameter of 49 μm and an analysis bandwidth of 4–9 MHz were chosen because they allowed the *ka* value (acoustic wave number times average scatterer radius) to be less than 1.2 over the analysis bandwidth. Insana *et al.*[6,7] showed that by keeping the *ka* value less than 1.2 the most accurate estimations of scatterer diameter could be obtained. The scattered pulses were then combined to create a single backscattered wavetrain. The time location of each scattered pulse depended on the location of the scatterer from the idealized source. In the attenuating medium simulation, the same matrix was used to construct the backscattered wave train except the scattered pulses were also reduced by the frequency-dependent attenuation corresponding to the distance of the scatterer from the source. A total of 25 random matrices were used for both the attenuated simulation and unattenuated simulation in order to obtain an average for the normalized power spectrum. Figure 4 shows a picture of a backscattered signal from an attenuated and an unattenuated simulated medium.

Estimates of average effective scatterer size were obtained by minimizing the average squared difference (MASD) between the normalized power spectrum obtained

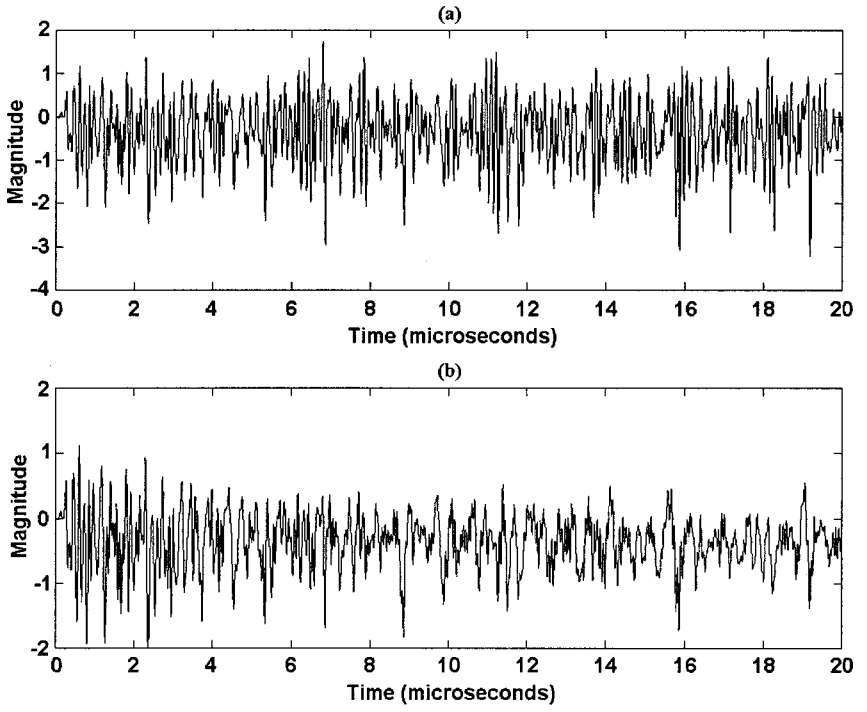


FIG. 4. Backscattered echos from a simulated medium with random scatterers when frequency-dependent attenuation is not present and attenuation is present; (a) unattenuated wave train; (b) attenuated wave train.

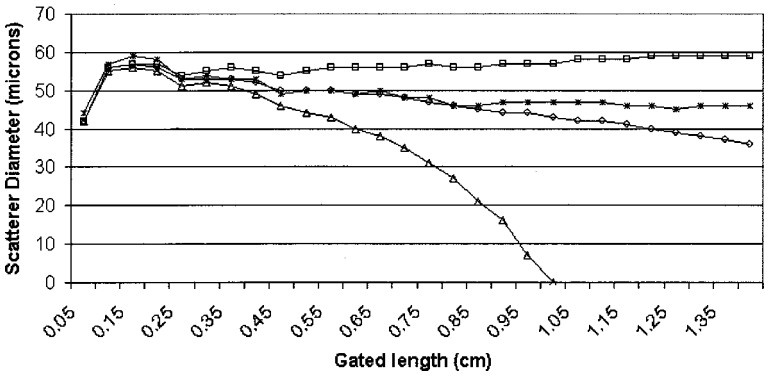


FIG. 5. Scatterer diameter estimations from simulated backscatter returns of randomly distributed Gaussian scatterers with average diameters of 49 micrometers using different attenuation-compensation algorithms and an unattenuated backscatter return, $\alpha = 7 \times 10^{-9} f^{1.19}$ Np/cm; (△) point compensation; (□) O'Donnell and Miller compensation; (◇) Oelze and O'Brien compensation; (*) unattenuated estimates.

from gated segments of the backscattered signal and the Gaussian form factor.[6] The MASD is defined as

$$\mathrm{MASD} = \min\left(\frac{1}{m} \sum_{i=1}^{m} (X_i - \overline{X})^2 \right) \tag{36}$$

with

$$X_i = 10 \log[W_{\mathrm{meas}}(f_i) / f_i^4 F_{\mathrm{Gauss}}(a_{\mathrm{eff}}, f_i)] \tag{37}$$

and

$$\overline{X} = \frac{1}{m} \sum_{i=1}^{m} X_i , \tag{38}$$

where m is the number of points in the frequency data bandwidth. The estimate of the average scatterer diameter is the argument that minimizes Eq. (36). The simulations were constructed so that there was no attenuation in the intervening medium ($e^{-2\alpha_0(f)x_0} = 1$). Estimates obtained from the unattenuated signal were used as the standard or control measurement. Estimates were then calculated from the attenuated signal without compensating for attenuation loss and by using the different compensation algorithms. Figure 5 shows the average scatterer diameters estimated from the unattenuated signal and attenuated signals compensated with point attenuation-compensation, the O'Donnell and Miller compensation function, and the Oelze and O'Brien compensation function. The frequency-dependent attenuation used in the simulation is a power law. The Oelze and O'Brien compensation algorithm gives the closest estimate of average scatterer diameter to the estimate by the unattenuated signal. As $\alpha(f)L$ continues to increase, all of the attenuation-compensation functions give estimates of scatterer diameter that diverge from the unattenuated case. The point attenuation and the Oelze and O'Brien attenuation-compensation functions cause an underestimation of the average scatterer diameter while the O'Donnell and Miller attenuation-compensation scheme gives an overestimation of the average scatterer diameter. This means that there exists some optimal attenuation-compensation routine between the Oelze and O'Brien compensation function and the O'Donnell and Miller compensation function.

The percent differences between the scatterer size estimates from the unattenuated signal and the attenuated signal, with and without compensation, are shown in Figs. 6 and 7, respectively. The attenuation used in Fig. 6 is much less than the attenuation used in Fig. 7. The attenuation values were chosen to represent possible values observed for soft tissues.[4,19,20] A comparison between Figs. 6 and 7 shows that with the higher attenuation, the divergence between the control measurement and the compensated measurements occurs at much shorter gate lengths. For example, the percent difference between the unattenuated estimates and the O'Donnell and Miller compensated estimates differ by 10% at a gate length of 7–8 mm for the smaller attenuation (Fig. 6) and at a gate length of about 4.5 mm for the higher attenuation (Fig. 7).

A further examination of Figs. 6 and 7 shows that not compensating for the effects of frequency-dependent attenuation on the power spectrum have immediate consequences on the estimation of average scatterer diameter. Not compensating the normalized power spectrum leads to poor estimates of scatterer sizes. For the lower attenuation case, the point compensation term gives good results for average scatterer size when the gate length is less than 1.25 cm. For the higher attenuation case, the point attenuation gives very poor

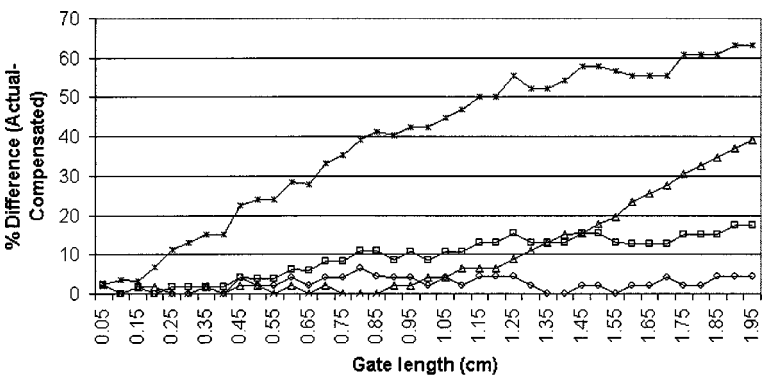


FIG. 6. Comparison of the present difference between the estimation of average scatterer diameter between the unattenuated scatter returns and attenuated scatter returns with various compensation routines, $\alpha = 7 \times 10^{-9} f^{1.13}$ Np/cm; (△) point compensation; (□) O'Donnell and Miller compensation; (◇) Oelze and O'Brien compensation; (*) uncompensated.

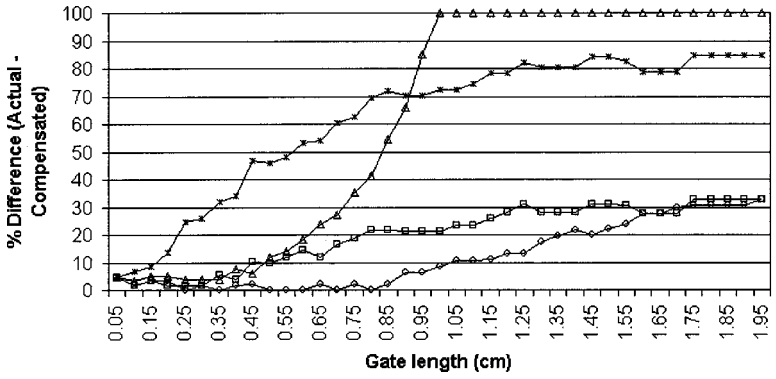


FIG. 7. Comparison of the percent difference between the estimation of average scatterer diameter between the unattenuated scatter returns and attenuated scatter returns with various compensation routines, $\alpha = 7 \times 10^{-9} f^{1.19}$ Np/cm; (△) point compensation; (□) O'Donnell and Miller compensation; (◇) Oelze and O'Brien compensation; (*) uncompensated.

estimates after a gate length of only 4 mm. The estimation by the point attenuation diverges much more rapidly than the other compensation schemes especially for the larger attenuation.

For the lower attenuation case, the Oelze and O'Brien compensation function gives estimates that are smaller than 10% over all gate lengths from the unattenuated estimations while the O'Donnell and Miller scheme gives errors of almost 20% at the larger gate lengths. In the higher attenuation case with a gate length of 8 mm, the Oelze and O'Brien compensation function gives estimates that are more than 20% better than the O'Donnell and Miller compensation function. Increasing the accuracy of estimates of scatterer size by 20% could be significant to applications that utilize this information, i.e., diagnosing tissue disease from average cell size. Comparisons of the O'Donnell and Miller compensation function with the Oelze and O'Brien compensation function show that better estimates can be obtained with the Oelze and O'Brien compensation term when scattering follows the Gaussian form factor.

Simulations were also run in which the time signal was gated using a Hanning window. Estimates were made of the average scatterer diameter from the signal gated with a Hanning window using the Oelze and O'Brien attenuation-compensation function, the O'Donnell and Miller attenuation-compensation function, and the Oelze and O'Brien attenuation-compensation function adjusted for the Hanning window according to Eq. (34). Figure 8 shows the percent difference between the estimates made using the different attenuation-compensation functions. The adjusted Oelze and O'Brien attenuation-compensation function yielded superior estimates of the average scatterer diameter over the other attenuation-compensation functions.

IV. OPTIMIZATION FOR LARGE ATTENUATION

The expected variance of measurements of particle diameters for particles following the Gaussian form factor has been calculated.[21] The expected variance of the measured average scatterer diameter using the MASD scheme has been shown to be inversely proportional to the actual average scatterer diameter squared.[21] The measurement is less sensitive to smaller average scatterer diameters. In terms of the normalized power spectrum, the fact that the measurement is less sensitive to smaller scatterers means that the normalized power spectrum is changing more slowly with frequency for smaller scatterer sizes, i.e., the overall slope for Gaussian form factor is flatter as seen in Fig. 1. Better estimates are made when the variability of the measured spectrum with frequency is greater rather than smaller.

Undercompensating the attenuation effect by a certain amount will give a more rapidly decreasing overall slope for the normalized power spectrum and a larger estimation of the average scatterer diameter. Overcompensating the attenuation by the same amount will give a smaller estimation of the average diameter. The overcompensation leads to a larger variance in the measurement. What this means for attenuation compensation is that it is better to undercompensate for the attenuation than to overcompensate by the same amount.

Furthermore, attenuation usually increases with increasing frequency. As the attenuation or gate length increases, the signal to noise ratio in the higher frequency components of

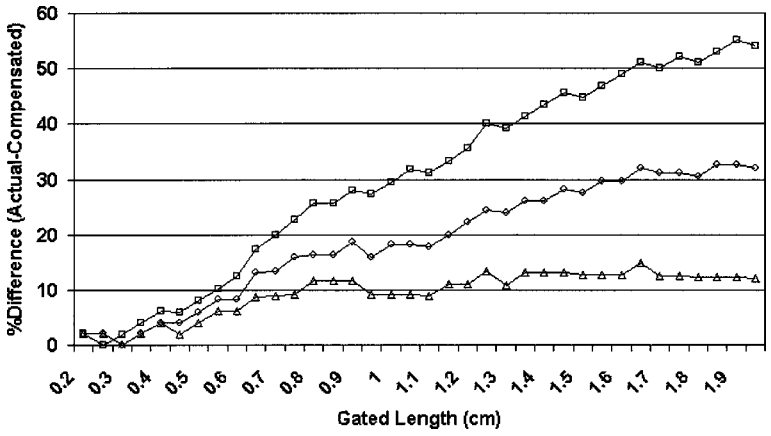


FIG. 8. Comparison of the percent difference between the estimation of average scatterer diameter between the unattenuated scatter returns and attenuated scatter returns using the Hanning window and various compensation routines, $\alpha = 8 \times 10^{-8} f$ Np/cm; (△) Oelze and O'Brien (Hanning adjusted) compensation; (□) O'Donnell and Miller compensation; (◇) Oelze and O'Brien compensation.

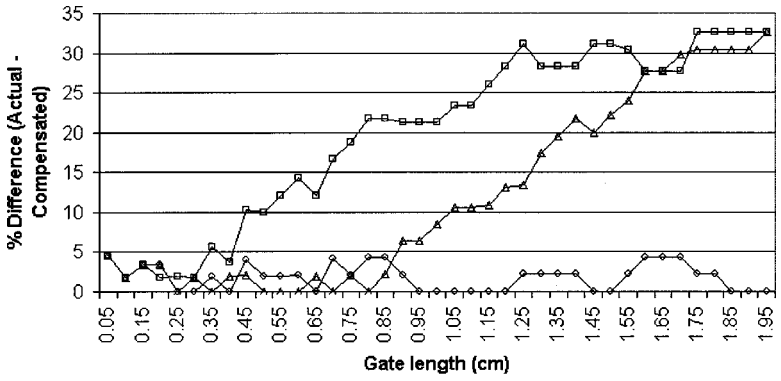


FIG. 9. Comparison of accuracy of different attenuation-compensation techniques vs increasing gate length with $\alpha = 7 \times 10^{-9} f^{1.19}$ Np/cm; (◇) combined compensation; (□) O'Donnell and Miller compensation; (△) Oelze and O'Brien compensation.

the normalized power spectrum becomes smaller. The attenuation-compensation functions compensate the attenuation losses but also amplify the noise. The amplification of noise, especially at high frequencies, tends to make the slope of the normalized power spectrum appear flatter and overcompensation occurs. When the signal to noise is small due to large attenuation, it is better to slightly undercompensate the normalized power spectrum for the attenuation losses. Evidence that undercompensation can yield better results when the signal is noisy can be seen from Figs. 7 and 5. From Fig. 5 it is seen that the O'Donnell and Miller compensation function yielded estimations that were larger in size than the actual (undercompensated) while the Oelze and O'Brien compensation function yielded smaller diameter estimations (overcompensated). Figure 7 shows that the O'Donnell and Miller compensation function, which undercompensates the attenuation losses, begins to give better results as the gate length and frequency-dependent attenuation losses become greater because of the noise effect.

Figure 2 shows the magnitude of the compensation terms versus increasing values of $\alpha(f)L$. From Fig. 2 it is seen that the O'Donnell and Miller compensation function give the smallest attenuation correction to the normalized power spectrum, point compensation gives the largest correction, and Oelze and O'Brien compensation lies somewhere in between. Even though the Oelze and O'Brien compensation function appears to yield a power spectrum that is closer to the unattenuated normalized power spectrum, with large enough $\alpha(f)L$ the O'Donnell and Miller compensation will yield better scatterer size estimations.

The Oelze and O'Brien compensation term leads to an underestimation of the average scatterer size with large $\alpha(f)L$ and the O'Donnell and Miller compensation gives an overestimation of the average scatterer diameter. By combining the two approaches, the estimation of average scatterer sizes can be optimized for large $\alpha(f)L$. At small to medium $\alpha(f)L, \alpha(f)L < 1$, the Oelze and O'Brien compensation function should be used and at larger $\alpha(f)L$ a weighted average of the Oelze and O'Brien compensation function and O'Donnell and Miller compensation function can be used. A closed-form compensation function accomplishing the combined approach can be obtained by use of an interpolation function, $I(\alpha(f)L)$, giving

$$A_{\text{comb}} = A_{\text{OO}} I(\alpha L) + [1 - I(\alpha L)](A_{\text{OO}} + 3A_{\text{OM}})/4, \quad (39)$$

where

$$I(\alpha L) = \frac{2}{1 + \exp(\alpha L/2)}.$$

In this particular case, a three to one weighting in favor of the O'Donnell and Miller attenuation-compensation function was found empirically to give the best estimates of scatterer diameter. As $\alpha(f)L$ becomes large, $\alpha(f)L > 1$, the attenuation compensation switches from Oelze and O'Brien compensation to a combined compensation involving both the Oelze and O'Brien compensation function and the O'Donnell and Miller compensation term. Estimations were then obtained with the combined approach using the same backscattered signals from Fig. 7. Figure 9 shows the percent difference between the compensation routines and the estimations obtained from the unattenuated signal. The combined approach continues to keep the accuracy of the Oelze and O'Brien compensation routine at smaller $\alpha(f)L$ and results in a marked improvement of scatterer estimations at large gate lengths ($\alpha(f)L > 1$). The combined approach gives estimates that are consistently less than 5% different from the estimates obtained form the unattenuated signal over all the gate lengths ($\alpha(f)L$) examined.

V. CONCLUSIONS

Compensating for the frequency-dependent attenuation loss of a signal in a medium is necessary to accurately estimate scatterer properties such as average scatterer diameter. Several methods exist for compensating the attenuation losses to the normalized power spectrum. The ability of the attenuation-compensation functions to accurately compensate for the attenuation losses depends on the relative size of $\alpha(f)L$. For small $\alpha(f)L, \alpha(f)L < 0.5$, most attenuation-compensation routines yield accurate results. When $\alpha(f)L$ becomes larger, some attenuation-compensation routines overcompensate leading to an underestimation of the scatterer sizes while other compensation routines undercompensate leading to an overestimation of scatterer sizes.

A new attenuation-compensation function was developed. The new attenuation-compensation function was then compared with other compensation functions through numerical simulations of random scattering media. The new attenuation-compensation function was found to give superior results to previous compensation routines except at very

large $\alpha(f)L, \alpha(f)L > 1$. At large $\alpha(f)L$ all attenuation-compensation routines lead to increasingly overestimation or underestimation of scatterer sizes.

When the attenuation is large, it is better to undercompensate the losses to the normalized power spectrum through frequency-dependent attenuation than to correctly compensate the attenuation losses because of the amplification of noise at the high frequencies. Analysis of the attenuation-compensation routines and their ability to properly estimate average scatterer diameter showed that the O'Donnell and Miller compensation function tended to give overestimates of the scatterer sizes. The Oelze and O'Brien attenuation-compensation routine gave an underestimation of scatterer size at large $\alpha(f)L$. The scattering amplitude at the higher frequencies will be smaller for the Gaussian form factor than at the lower frequencies. Attenuation causes the signal to noise ratio to decrease, especially at the higher frequencies. The amplification of noise at the high frequencies leads to an underestimation of the scatterer size from the compensated, normalized power spectrum.

An improvement to the overall attenuation compensation was achieved by combining two compensation functions through an interpolation function. The analysis indicates that the Oelze and O'Brien compensation routine yields a normalized power spectrum that is closest to the unattenuated case. However, because the O'Donnell and Miller compensation function undercompensates, then, as $\alpha(f)L$ becomes large enough, the O'Donnell and Miller compensation begins to give improved estimations over the Oelze and O'Brien compensation function. At lower values of $\alpha(f)L$, less than one, the attenuation compensation followed the Oelze and O'Brien compensation function. At larger values of $\alpha(f)L$, greater than one, the compensation was achieved through a weighted average between the O'Donnell and Miller and the Oelze and O'Brien compensation functions. The combined approach yielded estimates of scatterer size close to the unattenuated case for all values of $\alpha(f)L$ examined using the Gaussian form factor. If other form factors are used to describe scattering from random media, it may be that different combinations of the O'Donnell and Miller and the Oelze and O'Brien compensation functions would yield improved estimates.

The windowing function used to gate the rf time signal must also be accounted for to correctly compensate the attenuation losses to the normalized power spectrum. An adjustment to the Oelze and O'Brien attenuation-compensation function based on the type of windowing function can be easily implemented using Eq. (31). Accounting for the effects of the window on the normalized power spectrum has been shown necessary to accurately estimate scatterer properties.

In choosing an appropriate attenuation-compensation function, several factors need to be considered. The magnitude of the attenuation and the analysis frequency bandwidth must be taken into account. The attenuation may be small at low frequencies and relatively high at large frequencies. According to the simulations, the Oelze and O'Brien attenuation-compensation function will give better estimates for the smaller attenuation. However, if large attenuation of the higher frequencies in the analysis bandwidth leads to small signal to noise, then the Oelze and O'Brien attenuation compensation may not be the best choice for particular frequencies. If there is a small signal to noise ratio at the higher frequencies, then it is better to undercompensate those frequencies. Future work in this area would be to catalog the appropriate attenuation-compensation function or combination of functions for a particular gate length, window type, analysis bandwidth, form factor describing the scattering, frequency dependence, and magnitude of the attenuation.

ACKNOWLEDGMENT

This work was supported by NIH Grant Nos. CA09067 and CA79179.

[1] F. L. Lizzi, M. Greenbaum, E. J. Feleppa, and M. Elbaum, "Theoretical framework for spectrum analysis in ultrasonic tissue characterization," J. Acoust. Soc. Am. **73**, 1366–1371 (1983).
[2] F. L. Lizzi, M. Ostromogilsky, E. J. Feleppa, M. C. Rorke, and M. M. Yaremko, "Relationship of ultrasonic spectral parameters to features of tissue microstructure," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **34**, 319–329 (1987).
[3] J. A. Campbell and R. C. Waag, "Normalization of ultrasonic scattering measurements to obtain average differential scattering cross sections for tissues," J. Acoust. Soc. Am. **74**, 393–399 (1983).
[4] D. Nicholas, "Evaluation of backscattering coefficients for excised human tissues: Results, interpretation and associated measurements," Ultrasound Med. Biol. **8**, 17–22 (1979).
[5] D. K. Nassiri and C. R. Hill, "The use of angular scattering measurements to estimate structural parameters of human and animal tissues," J. Acoust. Soc. Am. **79**, 2048–2054 (1986).
[6] M. F. Insana, R. F. Wagner, D. G. Brown, and T. J. Hall, "Describing small-scale structure in random media using pulse-echo ultrasound," J. Acoust. Soc. Am. **87**, 179–192 (1990).
[7] M. F. Insana and T. J. Hall, "Parametric ultrasound imaging from backscatter coefficient measurements: Image formation and interpretation," Ultrason. Imaging **12**, 245–267 (1990).
[8] M. F. Insana, "Modeling acoustic backscatter from kidney microstructure using an anisotropic correlation function," J. Acoust. Soc. Am. **97**, 649–655 (1995).
[9] T. J. Hall, M. F. Insana, L. A. Harrison, and G. G. Cox, "Ultrasonic measurement of glomerular diameters in normal adult humans," Ultrasound Med. Biol. **22**, 987–997 (1996).
[10] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).
[11] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw–Hill, New York, 1968).
[12] F. L. Lizzi, M. Astor, T. Liu, C. Deng, D. J. Coleman, and R. H. Silverman, "Ultrasonic spectrum analysis for tissue assays and therapy evaluation," J. Imaging Syst. Technol. **8**, 3–10 (1997).
[13] V. Roberjot, S. L. Bridal, P. Laugier, and G. Berger, "Absolute backscatter coefficient over a wide range of frequencies in a tissue-mimicking phantom containing two populations of scatterers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 970–977 (1996).
[14] R. A. Sigelmann and J. M. Reid, "Analysis and measurement of ultrasound backscattering from an ensemble of scatterers excited by sine-wave bursts," J. Acoust. Soc. Am. **53**, 1351–1355 (1973).
[15] J. F. Chen, J. A. Zagzebski, and E. L. Madsen, "Tests of backscatter coefficient measurement using broadband pulses," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 603–607 (1993).
[16] M. O'Donnell and J. G. Miller, "Quantitative broadband ultrasonic backscatter: An approach to nondestructive evaluation in acoustically inhomogeneous materials," J. Appl. Phys. **52**, 1056–1065 (1981).
[17] K. A. Wear, R. F. Wagner, M. F. Insana, and T. J. Hall, "Application of autoregressive spectral analysis to cepstral estimation of mean scatterer spacing," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 50–58 (1993).

[18] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ, 1975).
[19] D. K. Nassiri, D. Nicholas, and C. A. Miles, ''Attenuation of ultrasound in skeletal muscle,'' Ultrasonics **17**, 230–232 (1979).
[20] K. A. Topp and W. D. O'Brien, Jr., ''Anisotropy of ultrasonic propagation and scattering properties in fresh rat skeletal muscle *in vitro*,'' J. Acoust. Soc. Am. **107**, 1027–1033 (2000).
[21] P. Chaturvedi and M. F. Insana, ''Error bounds on ultrasonic scatterer size estimation,'' J. Acoust. Soc. Am. **100**, 392–399 (1996).

Session 1aAO

Acoustical Oceanography and Underwater Acoustics: Ed Hamilton Memorial Session on Sediment Properties I

Mohsen Badiey, Chair
College of Marine Studies, University of Delaware, Robinson Hall, Newark, Delaware 19716

Chair's Introduction—9:30

Invited Papers

9:35

1aAO1. Attenuation and permeability: high frequency crosswell tomography for imaging permeability structure of sediments. Tokuo Yamamoto (AMP, RSMAS, Univ. of Miami, 4900 Rickenbaker Cswy., Miami, FL 33149), Junich Sakakibara (Kawasaki Steel Eng. & Construction Co., Tokyo, Japan), and Yoshiyuki Mohri (Natl. Inst. for Rural Eng., Tsukuba, Japan)

Imaging the permeability within sediments and rocks is one of the important problems left to advance in geoacoustics and geophysics. The Hamilton attenuation data suggests it is possible. As a U.S.–Japan joint research project we have selected this important problem. Desired permeability and porosity of test sediments are achieved with a sophisticated technique of sediment deposition called laminar sedimentation. To minimize reflections from the tank walls a wave absorber was placed between the tank wall and sediments. Placing a second hydrophone well between the source well and the receiver well, bore hole losses were physically minimized so as to assure accurate measurements of attenuation. A broadband piezoelectric source and hydrophone arrays were used to propagate and receive high frequency M-sequence acoustic waves. High resolution permeability images within various sediments are extracted from attenuation and velocity measured by high frequency (10 kHz to 80 kHz) acoustic cross-well tomography experiments and a Biot inversion. [Work supported by ONR, NIRE, and KSEC.]

9:55

1aAO2. Extending the legacy of Ed Hamilton in high frequency sediment acoustics. Eric I. Thorsos, Kevin L. Williams, Darrell R. Jackson, Dajun Tang (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), Michael D. Richardson, and Kevin B. Briggs (Naval Res. Lab., Stennis Space Center, MS 39529)

Ed Hamilton set a high standard for geoacoustic modeling of the sea floor. His comprehensive yet thorough descriptions of sediment properties in forms readily usable by acoustic modelers were seminal contributions to underwater acoustics. Hamilton was able to synthesize a wide diversity of *in situ* acoustic data within the framework of the Gassmann sediment model in which the pore water and sediment solids move together in response to acoustic fields and with attenuation modeled empirically. During the Sediment Acoustics Experiment in 1999 (SAX99), a wide variety of acoustics measurements were made in the 0.1–1000 kHz range combined with extensive sediment characterization of the medium sand, shallow water site. Some of our acoustic measurements support the hypothesis that relative motion between the pore fluid and sand grains has important consequences for sediment acoustics at the low end of our frequency range. Results from SAX99 will be reviewed which suggest the importance of more complex sediment models for a shallow water sand site. One such model will be described that recovers the spirit of Hamilton's desire to minimize complexity yet accurately describes the acoustic behavior of the sediment. [Work supported by ONR.]

10:15

1aAO3. Ocean sediment high-frequency acoustics. Nicholas P. Chotiros (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, chotiros@arlut.utexas.edu)

The upper few meters of ocean sediment is most relevant to acoustics, in the frequency band 10–500 kHz. It is often approximated as a stratified fluid or visco-elastic solid. The acoustics of the sediment, however, is often more complicated than either of these approximations will allow. A symptom of the problem is their inability to match experimental data. The problem seems to be centered about the physics of the grain–grain and grain–fluid interactions. Much of the difficulty lies in the physics of the vast range of structures that lie between the two extremes of a fluidized suspension, such as silt, and a fully cemented porous structure, such as sandstone. The Biot–Stoll equations of sound propagation in a porous medium provide the mathematical framework to represent the acoustics of a porous structure. The definition of pore fluid and skeletal frame is a critical issue that needs further development, in order to accommodate the full range of sediment types. [Work supported by ONR, Ocean Acoustics, Jeff Simmen.]

10:35–10:50 Break

Contributed Papers

10:50

1aAO4. An approximate Biot model for marine sediments with nonuniform pore sizes. Altan Turgut (Naval Res. Lab., Washington, DC 20375)

The issue of attenuation and dispersion in marine sediments has been investigated by Hamilton [Geophysics (1972)] in detail. In his paper, Hamilton stated that an adequate model for marine sediments must be anelastic in which viscous damping and velocity dispersion should be considered if the pore water moves significantly relative to mineral structure. He introduced an attenuation parameter k which is empirically related to mean grain size and porosity. Based on the experimental measurements in frequencies within 14–100 kHz, he suggested a nearly elastic model with linear frequency dependency of attenuation and negligible velocity dispersion. Then, he explained the peak value of k for fine sands under the context of frictional energy losses between grains. In this study, an approximate Biot model is proposed. The new model includes both frame anelasticity and viscous damping in marine sediments with nonuniform pore sizes. For most marine sediments, the approximate model predicts similar frequency dependency of attenuation and velocity dispersion of compressional and shear waves to those of predicted by the Biot model. It has been shown that mean grain size and porosity dependency of parameter k can be also predicted by the new model without introducing an empirical relation. [Work supported by ONR.]

11:05

1aAO5. Velocity dispersion in granular marine sediment. Masao Kimura (Dept. of Geo-Environ. Technol., Tokai Univ., 3-20-1 Orido, Shimizu, Shizuoka 424-8610, Japan, mkimura@scc.u-tokai.ac.jp)

Acoustic wave velocity dispersion in granular marine sediment is important to the study of sediment acoustics. The reported dispersion results cannot be explained with Biot's model using constant Biot parameters. In this study, the acoustic wave velocities in water-saturated glass beads and beach sands with different grain sizes are measured at different frequencies. The velocities in air-saturated same media and in same media in vacuum are also measured. Next the frame moduli for water-saturated media, air-saturated media, and media in vacuum are obtained from these measured velocities. From these results, it is seen that the frame bulk modulus has dependence on the grain size and the frequency. It is considered that one possibility of the velocity dispersion in water-saturated granular marine sediment is grain size and frequency-dependent frame bulk modulus due to the gap stiffness between the grains.

11:20

1aAO6. A comparison of models for water-saturated sand by inversion of reflection loss measurements. Marcia J. Isakson, Sam McJunkin, Egbert Hood III, and Nicholas Chotiros (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029, misakson@arlut.utexas.edu)

Our objective is an accurate model of sediment acoustics that is consistent with all measurable acoustic and geophysical properties. This will lead to a better understanding of acoustic penetration in ocean sediments. In this study, reflection data taken from a smooth water/sand interface are inverted using a broad range of models. The models considered in the inversion are the visco-elastic solid model, the Biot/Stoll model, the composite Biot model, the effective density model, and the Buckingham model. Inversions were performed using a simulated annealing optimization algorithm with OASR as the forward model. Each model is considered for its ability to produce realistic parameters and its fit to the experimental data. [Work supported by ONR, Undersea Signal Processing, John Tague.]

11:35

1aAO7. Determination of the absorbing and scattering properties of shallow-water sediment using wideband signals. Boris G. Katsnelson, Valery A. Grigoryev (Voronezh Univ., 1, Universitetskaya Sq., Voronezh 394693, Russia), and Valery G. Petnikov (Gen. Phys. Inst., Moscow 117333, Russia)

Sound propagation in a shallow-water region is studied for bottom parameter estimation within the framework of MFP (matched field processing) method. The experiment was conducted in the Barents Sea at the stationary acoustic track (about 14 km length). Basic bottom parameters are determined on the basis of comparison between experimental and theoretical frequency spectra of a signal received from a wideband source (20–100 Hz). A shallow-water waveguide model is developed on the basis of measured bathimetry, sound-speed profile, and two-component model of the bottom sediment. Porosity and coefficients of absorption and volume scattering are considered as the parameters, characterizing bottom sediment. The connection between porosity and mentioned coefficients is taken from Hamilton's data [J. Acoust. Soc. Am. **59**, 528–535 (1976)]. Porosity (correspondingly, density and sound speed in sediment) is obtained, and a conclusion is made about relative contribution of different mechanisms of losses (absorption and volume scattering) in the seafloor at different sound frequencies. [Work supported by CRDF and ME RF, Award No. VZ-010-0.]

Session 1pAA

Architectural Acoustics: Auditorium and Concert Hall Acoustics

David E. Marsh, Chair
Pelton Marsh Kinsella, 1420 West Mockingbird Lane, Suite 400, Dallas, Texas 75247-4932

Contributed Papers

1:30

1pAA1. Effects of sound source directivity on auralizations. Nathan W. Sheets and Lily M. Wang (Architectural Eng. Prog., Univ. of Nebraska, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, lwang@unl.edu)

Auralization, the process of rendering audible the sound field in a simulated space, is a useful tool in the design of acoustically sensitive spaces. The auralization depends on the calculation of an impulse response between a source and a receiver which have certain directional behavior. Many auralizations created to date have used omnidirectional sources; the effects of source directivity on auralizations is a relatively unexplored area. To examine if and how the directivity of a sound source affects the acoustical results obtained from a room, we used directivity data for three sources in a room acoustic modeling program called Odeon. The three sources are: violin, piano, and human voice. The results from using directional data are compared to those obtained using omnidirectional source behavior, both through objective measure calculations and subjective listening tests.

1:45

1pAA2. Acoustic radiosity for computation of sound fields in diffuse environments. Ralph T. Muehleisen and C. Walter Beamer IV (Dept. of Civil, Environ., and Architectural Eng., Univ. of Colorado, Boulder, CO 80309)

The use of image and ray tracing methods (and variations thereof) for the computation of sound fields in rooms is relatively well developed. In their regime of validity, both methods work well for prediction in rooms with small amounts of diffraction and mostly specular reflection at the walls. While extensions to the method to include diffuse reflections and diffraction have been made, they are limited at best. In the fields of illumination and computer graphics the ray tracing and image methods are joined by another method called luminous radiative transfer or radiosity. In radiosity, an energy balance between surfaces is computed assuming diffuse reflection at the reflective surfaces. Because the interaction between surfaces is constant, much of the computation required for sound field prediction with multiple or moving source and receiver positions can be reduced. In acoustics the radiosity method has had little attention because of the problems of diffraction and specular reflection. The utility of radiosity in acoustics and an approach to a useful development of the method for acoustics will be presented. The method looks especially useful for sound level prediction in industrial and office environments. [Work supported by NSF.]

2:00

1pAA3. Pilot study: Exposure and materiality of the secondary room and its impact in the impulse response of coupled-volume concert halls. Michael Ermann and Marty E. Johnson (Virginia Tech Dept. of Architecture, 201 Cowgill Hall 0205, Blacksburg, VA 24061, mermann@vt.edu)

What does one room sound like when it is partially exposed to another (acoustically coupled)? More specifically, this research aims to quantify how operational and design decisions impact aural impressions in the design of concert halls with acoustical coupling. By adding a second room to a concert hall, and designing doors to control the sonic transparency between the two rooms, designers can create a new, coupled acoustic. Concert halls use coupling to achieve a variable, longer, and distinct reverberant quality for their musicians and listeners. For this study, a coupled-volume shoebox concert hall was conceived with a fixed geometric volume, form, and primary-room sound absorption. Aperture size and secondary-room sound-absorption levels were established as variables. Statistical analysis of sound decay in this simulated hall suggests a highly sensitive relationship between the double-sloped condition and (1) Architectural composition, as defined by the aperture size exposing the chamber and (2) Materiality, as defined by the sound absorbance in the coupled volume. Preliminary calculations indicate that the double-sloped sound decay condition only appears when the total aperture area is less than 1.5% of the total shoebox surface area and the average absorption coefficient of the coupled volume is less than 0.07.

2:15

1pAA4. Computer modeling of loudspeaker arrays in rooms. Roger Schwenke (Meyer Sound Labs., 2832 San Pablo Ave., Berkeley, CA 94702-2204, rogers@meyersound.com)

Loudspeakers present a special challenge to computational modeling of rooms. When modeling a collection of noncorrelated sound sources, such as a group of musicians, coarse resolution power spectrum and directivities are sufficient. In contrast, a typical loudspeaker array consists of many speakers driven with the same signal, and are therefore almost completely correlated. This can lead to a quite complicated, but stable, pattern of spatial nulls and lobes which depends sensitively on frequency. It has been shown that, to model these interactions accurately, one must have loudspeaker data with 1 deg spatial resolution, 1/24 octave frequency resolution including phase. It will be shown that computer models at such a high resolution can in fact inform design decisions of loudspeaker arrays.

2:30

1pAA5. Under-balcony acoustics in concert halls: Vertical energy loss at the listener and acceptable view angles deduced from its effects. Young-Min Kwon and Yasushi Shimizu (Architectural Acoust., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180)

Recent investigations into acoustics in under-balcony spaces have found that balcony overhangs reduce the early energy from above as well as the late energy from the main space, and hence cause an acoustic defect in the vertical. Based upon these findings, this research analyzes energy loss under the balcony, when sound energy comes from above through a vertical median plane. Fundamentally, it aims to find acoustically acceptable view angles (Θ) against the balcony opening height (H). Scale-model experiments in an anechoic room were done to obtain the amount of energy reduction (or reduction coefficient) by balcony overhangs as a function of incidence angle. Next, psychoacoustical experiments associated with listener envelopment were conducted under synthetic sound fields with constant lateral energy but with variable vertical energy to determine the appropriate Θ. These experiments, which combined with data from the scale-model experiments, cover paired comparison tests un-

der two sets: with balcony and without balcony conditions. Results of the scale-model experiments show that sound energy decreases greatly within certain incidence angles near the balcony and at lower frequency bands. The minimum vertical angle of view is inferred to be 30–40 deg from psychoacoustical experiments. Overhang depth (D), D/H ratio, and their relationship are also discussed.

2:45–3:00 Break

3:00

1pAA6. Evaluation of stage acoustics in a concert hall by measuring stage support. Jin Yong Jeon, Byung Kwon Lee (School of Architectural Eng., Hanyang Univ., Seoul 133-791, Korea, jyjeon@hanyang.ac.kr), and Michael Barron (Univ. of Bath, Bath, Somerset BA2 7AY, UK)

Although stage acoustics is a very important characteristic in a concert hall in regard to the acoustical quality on stage and in the audience, it is the fact that there has not been much academic concern about stage acoustics. In this study, stage acoustics was evaluated by using real, scale and computer models of Seoul Arts Center Concert Hall. The stage acoustics were measured and evaluated at eight positions on the stage by ST1 (Stage Support) initiated by Gade [Acustica **65**, 193–203 (1989)] as an early support from stage. The effect of diffusers on stage walls, the geometrical shape of the stage and the arrangement of performers on stage were investigated. In the results, it was found that the closer the receiving positions to stage walls, the higher the ST1 value, especially in the range of 3–5 m. The height of the stage ceiling or the ceiling reflector is also an influential aspect on stage support; the optimum height of the reflector started from 8 m.

3:15

1pAA7. Bayesian decay time analysis in coupled spaces using a proper decay model. Ning Xiang (Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677) and Paul M. Goggans (Univ. of Mississippi, Anderson Hall, University, MS 38677)

Acoustically coupled spaces have recently been drawing more and more attention in the architectural acoustics community. Determination of decay times in these coupled spaces from measured Schroeders decay function has become increasingly significant. In architectural acoustics practice, identification of decay times from measurements of room impulse responses in these coupled spaces can often require considerable effort. Recent works [N. Xiang, P. M. Goggans, and D. Li, J. Acoust. Soc. Am. **109**, 2283 (2001); P. M. Goggans and N. Xiang, *ibid.* **109**, 2383 (2001); N. Xiang and P. M. Goggans, *ibid.* **110**, 1415–1424 (2001)] have demonstrated that Bayesian parameter estimation and model selection prove to be powerful tools for evaluating decay times in coupled spaces. Following a brief introduction to Bayesian model selection, this paper will discuss further results from decay-time evaluations of measured room impulse responses. Emphasis will be given to the behavior of the decay model selection algorithm.

3:30

1pAA8. An investigation of platonic polyhedron loudspeaker systems as omnidirectional sources of sound. Timothy Leishman, Gordon Dix, Heather Smith, and Angela Tygerson (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602, tim_leishman@byu.edu)

For reasons of consistency, many room acoustics measurements (e.g., ISO 140-4 and ISO 3382) require the excitation of sound fields via omnidirectional sources. For years, acousticians have used dodecahedron loudspeaker enclosures with individual loudspeakers mounted on each face as conventional ''omnidirectional'' sources. However, frequency-dependent interference between loudspeakers, diffraction effects, and inherent directivities of individual loudspeakers must contribute to a departure from the omnidirectional ideal. As a result, criteria have been established to limit the maximum allowable deviations from omnidirectional radiation for qualified measurements. Despite such concessions, acousticians should consider an important question: is anything extraordinary about the dodecahedron loudspeaker configuration? The dodecahedron is merely one member of a family of five platonic polyhedrons. Other members of the family include the tetrahedron, hexahedron, octahedron, and icosahedron. How do the omnidirectional characteristics of loudspeaker systems based on these latter geometries compare to those based on the dodecahedron? This paper discusses several theoretical, numerical, and experimental results that provide answers to this question.

3:45

1pAA9. Characterization of the acoustic response of an auditorium for its recreation in a rehearsal room. Paolo Martignon (Phys. Dept., Univ. of Parma, Italy), Michele Zanolin (MIT, Boston, MA), and Stefano De Stabile (DEXT S.r.L., Parma, Italy)

A characterization of the stage acoustic response of the auditorium Niccolò Paganini in Parma (Italy) was carried out. The aim is to reproduce, in a low reverberation room, the response of this auditorium by means of a loudspeakers array system, implementing the image sources corresponding to first reflections and the correct isotropy of the reverberant field [A. J. Berkhout, J. Audio Eng. Soc. **36**, 977–995 (1988)]. The original idea is that such a room would become a virtual stage for orchestra rehearsals. Three linear microphone arrays 10 cm apart, arranged as Cartesian axes, were employed to acquire three impulse response arrays using one omnidirectional source. Measurement bandwidth was 20–2000 Hz: within this frequency range and the angles involved, spatial aliasing due to microphone array spacing (10 cm) does not occur (microphone spacing is a little less than the one of speakers array to be employed for reproduction). Data processing gave information about first reflections wave fronts curvature, and consequently their correspondence with architectural elements; after identifying these elements as different contributions, spectral absorption of each was calculated. Reverberation was separately measured in several points with different directions, using a directional microphone (W. P. Bruijn *et al.*, 104th AES Conv., preprint 4690, 1998).

4:00

1pAA10. An experimental and theoretical investigation of the high frequency broadband soundfield. Linda Franzoni (Dept. of Mech. Eng., Duke Univ., Durham, NC 27708-0300)

A relatively simple theoretical model of a rectangular enclosure with sound sources on one wall, an absorptive end-wall, and two absorptive side-walls is developed. Using a power balance at cross sections, a differential equation for mean-square pressure in the acoustic enclosure is derived. A simple formula for the mean-square pressure as a function of distance from source-wall to absorbing end-wall is obtained. The other variables in the formula include the random incidence absorption coefficients, the cross-sectional area and perimeter, and the source power. This formula is compared to results from numerical simulations, to a revised Sabine prediction for mean-square level, and to experimental results. For the most absorptive cases, however, a more refined theoretical model is needed in order to have reasonable agreement with the experimental data. In the refined model, it is assumed sources are distributed on an end-wall and the resulting sound waves are considered on an angle-by-angle basis. The mean-square pressure is calculated by integrating over all possible angles. Although still a fairly simple result, the effects of angle of incidence on the absorption and the reflection field in the enclosure are retained.

Session 1pAO

Acoustical Oceanography and Underwater Acoustics: Ed Hamilton Memorial Session on Sediment Properties II

Mohsen Badiey, Chair
College of Marine Studies, University of Delaware, Robinson Hall, Newark, Delaware 19716

Chair's Introduction—1:30

Invited Papers

1:35

1pAO1. Shear waves in surficial sediments: Revisited. Michael D. Richardson (Naval Res. Lab., Stennis Space Center, MS 39529-5004)

Numerous measurements of compressional and shear wave speed and attenuation have been made since Ed Hamilton's classic works on sediment geoacoustic properties (1955–1987). Recent measurements of shear wave speed and attenuation are summarized; regressions between shear wave speed and attenuation and easily measured sediment physical properties are used to demonstrate the predictability of these geoacoustic properties. The effect of overburden pressure (effective stress) on shear wave speed is also discussed. [Work supported by ONR.]

1:55

1pAO2. Empirical relationships among sediment properties and acoustic properties of the continental shelf seafloor. Kevin Briggs (Seafloor Sci. Branch, Naval Res. Lab., Stennis Space Center, MS 39529, kbriggs@apl.washington.edu)

An extensive database of sediment properties has been developed from which empirical relationships are made among geological index properties and geoacoustic properties. All data are derived from surficial shelf sediments carefully collected by divers or by subcoring from undisturbed box cores. The chief problem encountered in regressing one parameter on another is due to the variability inherent in marine sediments. The variability may be due to vertical gradients as well as sediment sources. By using various statistical techniques including multiple regression, sediment sound speed, sediment sound attenuation, and sediment density are related to sediment grain-size parameters of mean grain size, grain sorting, percent gravel, percent sand, percent silt, and percent clay. From these empirical fits one can predict acoustic properties from sediment grain size with varying levels of confidence. [Work supported by ONR.]

Contributed Papers

2:15

1pAO3. Inversion of shallow water ocean acoustic data using a complex plane-wave reflection coefficient approach. Steven Stotts, David Knobles, Robert Koch, Adam Cook (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029), David Grant, and Karl Focke (Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

A previously reported method of inversion in range-dependent environments using a complex plane-wave approach [Stotts *et al.*, J. Acoust. Soc. Am. **109**, 2334 (2001)] has been applied to data analysis. The method separates the propagation in the water column from the interaction with the seabed by first calculating and then storing the eigenray characteristics associated with the assumed known waveguide geometry. The inversion is then performed by calculating the plane-wave reflection coefficients at the predetermined angles and constructing the modeled field of the entire propagation for each perturbation of the seabed parameters. Results of applications of this model to data taken off the coast of Florida during the spring of 2001, will be presented. Consistent inversion results obtained from other models along with an examination of received time series, as well as sound speed profiles collected over the region, will also be discussed in the context of range-dependent inversion analysis. Comparisons of this method to a direct reflection coefficient inversion approach provide insight into the use of reflection coefficients as an inversion product.

2:30–2:45 Break

2:45

1pAO4. Rapid geo-acoustic characterization from a seismic survey. Kevin D. Heaney (ORINCON Corp., 4350 N. Fairfax Dr., Arlington, VA 22203), Daniel Sternlicht, Arthur Teranishi, Brett Castille, and Michael Hamilton (ORINCON Corp., 9363 Towne Ctr. Dr., San Diego, CA 92121)

A recent transmission loss experiment was conducted in Long Beach Harbor for the THUMS Long Beach Company. The objective of the experiment was to measure the range at which the received level was 160 dB for compliance with Marine Mammal regulations. This short experiment provided the opportunity to test the rapid geo-acoustic characterization (RGC) algorithm and perform real-time geo-acoustic inversions from a seismic source. The airgun source transmitted pulses every 20 s corresponding to every 45 m. The water depth was 10–15 m and the water was assumed to be iso-velocity. The data quality was excellent, providing clear striation patterns in the broadband frequency display. The RGC algorithm matches the observed time-spread, striation slope, and TL slope to precomputed values using a normal mode algorithm and parametric geoacoustic profiles based on Hamilton and Bachman's model. Precomputation of the acoustic observables, combined with real-time signal processing permits real time geo-acoustic characterization.

3:00

1pAO5. Rapid geo-acoustic characterization from a surface ship of opportunity. Kevin D. Heaney (ORINCON Corp., 4350 N. Fairfax Dr., Arlington, VA 22203)

The recent shift from deep water antisubmarine warfare to the littoral has brought about a dramatic change in the requirements of our understanding of the acoustic environment. In particular, shallow water acoustic propagation is dominated by the interaction of sound with the bottom and is therefore very sensitive to the geo-acoustic parameters of the local seafloor. The rapid geo-acoustic characterization (RGC) algorithm is presented, that determines a geo-acoustic model that best matches the general acoustic propagation parameters. Using data from a passing surface ship, the time spread, TL versus range and striation slope are measured at various frequencies. These are then matched to a simple two-layer geo-acoustic model based on the empirical curves of Hamilton and Bachman. The resulting estimate does reproduce the global features of the acoustic field that are relevant to acoustic prediction. This method is robust, rapid, and produces the relevant acoustic predictions.

3:15–3:35
Panel Discussion

MONDAY AFTERNOON, 3 JUNE 2002 RIVERS ROOM, 1:55 TO 4:00 P.M.

Session 1pMU

Musical Acoustics and Speech Communication: Naturalness in Synthesized Speech and Music

Sten O. Ternstrom, Chair
Department of Speech, Music & Hearing, Kungliga Tekniska Hogskolan, Drottning Kristinas Vag 31, Stockholm SE 10044, Sweden

Chair's Introduction—1:55

Invited Papers

2:00

1pMU1. Naturalness of voice qualities associated with adjustments in the glottis and lower vocal tract. Ingo R. Titze (Dept. of Speech Pathol. and Audiol., Univ. of Iowa and Wilbur James Gould Voice Ctr., Denver Ctr. for the Performing Arts, Denver, CO 80204, ingo-titze@dcpa.org)

Voice qualities such as twang, yawn, ring, pressed, sigh, and creaky have recently been captured by physical modeling of glottal aerodynamics, vocal fold biomechanics, and vocal tract acoustics. This paper focuses on the interactions between source and vocal tract adjustments and their likely co-variations in human voice production. For example, is pressed voice more likely to occur with twang than yawn, and is sigh (of fluty voice) more likely to occur with yawn than twang or ring? The underlying principles may be impedance matching and maximum source-filter interaction. [Work supported by NIH.]

2:30

1pMU2. Systematic downgrading for investigating "naturalness" in synthesized singing using STRAIGHT: A high quality VOCODER. Hideki Kawahara (Faculty of Systems Eng., Wakayama Univ., 930 Sakaedani, Wakayama 640-8510, Japan)

Conceptual simplicity of the classical channel VOCODER provides a powerful means for systematic investigations on perceptual effects of speech related physical parameters when combined with modern computational power and signal processing theories. A modern version of channel VOCODER, STRAIGHT [Kawahara *et al.*, Speech Commun. **27**, 187–207 (1999)], which is also an extension to pitch-synchronous analysis and synthesis, generates naturally sounding resynthesized speech from the analyzed smooth time-frequency surface and source parameters such as $F0$. This high-quality resynthesis enables close investigations on naturalness deterioration as a function of feature modifications in the decomposed parameter domain; for example, detailed shape of a $F0$ trajectory, underlying parameters to determine $F0$ trajectory dynamics, group delay alignment of excitation pulses and aperiodicity/periodicity ratio of the excitation source and so on. One of potential advantages of this strategy is based on the fact that our perceptual function is highly nonlinear. The other source of advantage is virtually an independent parameter set which allows precise control of parameter deviations from the original analysis results. An overview of recent findings and modification demonstrations will be presented. [Work supported by CREST grant of Japanese Science and Technology Corporation.]

3:00

1pMU3. Trumpet synthesis using context-dependent envelopes. Roger B. Dannenberg (School of Computer Sci., Carnegie Mellon Univ., Pittsburgh, PA 15213)

Synthesizing trumpet music in a natural-sounding way requires careful control. Even when synthesis is achieved by splicing together actual recorded trumpet tones, the result can sound artificial and unnatural. This is because natural notes are not played in isolation and are therefore influenced by neighboring notes and the musical context. In fact, a succession of notes played on the trumpet is likely to be a continuous sound with no separating silences. Improved synthesis can be obtained by calculating amplitude

and frequency control envelopes that take context into consideration. In the combined spectral interpolation synthesis (CSIS) method, the spectrum is controlled by instantaneous frequency and rms amplitude. These, in turn, are controlled by envelopes computed by a rule-based system. To reduce the high dimensionality of envelopes (typically a vector of 200 samples per second), envelopes are specified by about ten parameters. This reduced set of parameters is computed from note attributes, most importantly, the duration and pitches of the current and preceding notes, and whether or not the notes are tongued or slurred. This procedure is described in detail, and synthesis results will be demonstrated.

3:30

1pMU4. Spectral models for musical sound synthesis: Expression issues. Xavier Serra (Audiovisual Inst., Pompeu Fabra Univ., Passeig de Circumvallacio, 8. 08003 Barcelona, Spain)

In order to obtain a naturally sounding musical instrument or voice, we require sound models that can both capture the microstructure characteristics of the acoustical source and that at the same time can be controlled by meaningful parameters, in a similar way as the performers do. The issue of how to generate these control parameters is one of the most interesting and open problems in the area of musical sound synthesis. This presentation will focus on the control question in the context of two recently developed synthesizers based on a spectral modeling approach: a singing voice synthesizer and a wind instrument synthesizer. The specific spectral models used in each case will be presented followed by a discussion on the expression controls employed and how they are obtained from actual performances. These controls are mainly at the note and the note-to-note articulation level and affect parameters such as pitch, dynamics, brightness, and vibrato. From actual performances, templates and models of the temporal evolution of these parameters are extracted and then used as the basic controls of the synthesizer on top of which the user interacts in real time.

MONDAY AFTERNOON, 3 JUNE 2002 DUQUESNE ROOM, 1:30 TO 3:05 P.M.

Session 1pNS

Noise: Impulse and Continuous Noise-Effects, and Source Identification

John P. Seiler, Chair

U.S. DOL-Mine Safety and Health Administration, Cochran Mill Road, Building 038, Pittsburgh, Pennsylvania 15236

Chair's Introduction—1:30

Contributed Papers

1:35

1pNS1. Effects of acoustic impulses on hearing. Gerald Fleischer, Reinhard Müller, Guido Heppelmann, and Thomas Bache (Dept. of Auditory Res., Univ. of Giessen, Aulweg 123, D-35392 Giessen, Germany)

It is well known that acoustic impulses are especially dangerous to the ear. In order to understand the damaging mechanisms involved, cases of acute acoustic trauma in man were systematically collected and documented for many years. When possible, the damaging impulses were recreated and measured, to correlate the impulses with the auditory damage they caused. Detailed pure-tone audiometry up to 16 kHz was used to determine the effects on hearing. Together with epidemiological studies on various occupations, three different damaging mechanisms can be discerned. Relatively long and massive impulses (some explosions, some airbags) often lead to damage at low frequencies, from about 0.5 to 1.5 kHz. The typical notch at about 4 to 6 kHz typically is the result of strong peaks, lasting several milliseconds, or longer. There is another notch at 12 to 14 kHz, characteristic of very short, needle-like impulses that are caused by many hand weapons, toy pistols, and firecrackers. Probable mechanisms are discussed.

1:50

1pNS2. Infrasounds and biorhythms of the human brain. Ryszard Panuszka, Zbigniew Damijan, Cezary Kasprzak (Structural Acoust. and Biomed. Eng. Lab., Staszic Univ. of Cracow, AGH, Poland, ghpanusz@cyf-kr.edu.pl), and James McGlothlin (Health Sci. Dept., Purdue Univ., West Lafayette, IN)

Low Frequency Noise (LFN) and infrasound has begun a new public health hazard. Evaluations of annoyance of (LFN) on human occupational health were based on standards where reactions of human auditory system and vibrations of parts of human body were small. Significant sensitivity has been observed on the central nervous system from infrasonic waves especially below 10 Hz. Observed follow-up effects in the brain gives incentive to study the relationship between parameters of waves and reactions obtained of biorhythms (EEG) and heart action (EKG). New results show the impact of LFN on the electrical potentials of the brain are dependent on the pressure waves on the human body. Electrical activity of circulatory system was also affected. Signals recorded in industrial workplaces were duplicated by loudspeakers and used to record data from a typical LFN spectra with 5 and 7 Hz in a laboratory chamber. External noise, electromagnetic fields, temperature, dust, and other elements were controlled. Results show not only a follow-up effect in the brain but also a

result similar to arrhythmia in the heart. Relaxations effects were observed of people impacted by waves generated from natural sources such as streams and waterfalls.

2:05

1pNS3. Performance of electroacoustic hearing protectors. William Murphy and Mark Little (NIOSH, Hearing Loss Prevention Section, 4676 Columbia Pkwy., M.S. C-27, Cincinnati, OH 45226-1998, wjm4@cdc.gov)

The spectral and temporal response characteristics of five models of electroacoustic earmuffs (Bilsom 707 Impact II, Peltor Tactical 6s, Howard Leight Lightning and Thunder w/ProEars, and Silencio Electronic Low Pro hearing protectors) were measured with noise created by small arms fire from a 0.223 caliber rifle and a 9 mm handgun. Each hearing protector was tested in the fully-amplified (active) and no amplification (passive) conditions. The protectors were measured on an acoustic test fixture designed for high-level noise measurements. The protected and unprotected signals were recorded to digital audio tape and analyzed off-line. The reduction of peak sound pressure levels ranged between 16 and 32 dB for both the active and passive conditions. The time-averaged acoustic waveforms underneath the protectors exhibited little difference between active and passive conditions. The active electroacoustic performance approaches the passive attenuation performance. Auditory damage units were calculated for the unprotected and protected responses using the AHAAH cochlear model [Price and Kalb, J. Acoust. Soc. Am. **90**, 219–227 (1991)].

2:20

1pNS4. Problems associated with noise measurements in the mining industry. Eric R. Bauer (NIOSH, Pittsburgh Res. Lab., P.O. Box 18070, 626 Cochrans Mill Rd., Pittsburgh, PA 15236) and Jeffrey S. Vipperman (Univ. of Pittsburgh, Pittsburgh, PA 15261)

In response to the continuing problem of noise-induced hearing loss (NIHL) among mine workers, the National Institute for Occupational Safety and Health (NIOSH) has been conducting numerous noise- and hearing-loss research efforts in the mining industry. Research is underway to determine worker noise exposure, equipment noise, hearing loss and hearing protection use, and to evaluate engineering controls. Issues that are peculiar to the mining industry have complicated these efforts. A few of the issues that must be overcome to conduct meaningful research include constantly moving equipment, changing work environments, confined space, varying production rates, multiple noise sources, and electronic permissibility of instrumentation. This presentation will address the factors that affect the measurement and analysis of noise in the mining industry and how these factors are managed. In addition, some examples of research results will be included.

2:35

1pNS5. Dragline noise survey. Jeffrey S. Vipperman (Dept. of Mech. Eng., Univ. of Pittsburgh, 531 Benedum Hall, Pittsburgh, PA 15228) and Eric R. Bauer (Natl. Inst. for Occupational Safety and Health, Pittsburgh Res. Lab., Pittsburgh, PA 15236)

It is estimated that 70%–90% of miners have enough noise induced hearing loss (NIHL) to be classified as a disability (NIOSH, Publication No. 76-172, 1976; Franks, NIOSH Internal Report, 1996). In response, NIOSH is conducting a cross-sectional survey of the mining industry in order to determine the sources of mining noise and offer recommendations on how to mitigate high noise levels, and bring mining operations into compliance with the recent mining noise regulation: 30CFR, Part 62. This paper will outline the results from noise surveys of eight draglines which operate in above-ground coal mining operations. The data recorded include noise dosimetry in conjunction with time-at-task studies and 1/3-octave sound level (L_{eq}, L_{min}, and L_{max}) measurements. The 1/3-octave band readings were used to create noise contour maps which allowed the spatial and frequency information of the noise to be considered. Comparison of L_{min} and L_{max} levels offer insight into the variability of the noise levels inside the dragline. The potential for administrative controls is limited due to consistently high noise levels throughout the deck. Implementation of engineering controls is also hindered by the size and number of the noise sources and the frequency content of the noise.

2:50

1pNS6. Identification and classification of noise sources in a chain conveyor. John P. Homer (Dept. of Mech. Eng., Univ. of Pittsburgh, 531 Benedum Hall, Pittsburgh, PA 15228), Jeffrey S. Vipperman (Univ. of Pittsburgh, Pittsburgh, PA 15228), and Efrem R. Reeves (Natl. Inst. for Occupational Safety and Health, Pittsburgh Res. Lab., Pittsburgh, PA 15236)

Noise induced hearing loss (NIHL) is one of the most significant disabilities of workers in the mining industry. In response, the National Institute of Occupational Safety and Health (NIOSH) is conducting a study associated with mining equipment. This study outlines the analysis of a chain conveyor. Band-limited accelerometer, sound-intensity, far-field and near-field microphone measurements were taken along the conveyor section. The sound intensity measurements were used to identify areas with high noise as well as to calculate and 1/3-octave sound power levels. The total sound power results were used to classify the dominant noise sources where the 1/3-octave sound power results were used to identify the most contributive frequency bands to the overall noise of the system. Coherence analysis was performed between accelerometer and microphone measurements to identify structure–borne and air–borne noise paths of the system. Summary results from the analysis include recommendations for transmission control and damping devices and their ability to reduce noise to regulatory acceptable levels.

Session 1pPP

Psychological and Physiological Acoustics: Masking and Discrimination

Lawrence L. Feth, Chair
Speech and Hearing Science, The Ohio State University, 110 Pressey Hall, 1070 Carmack, Columbus, Ohio 43210-1372

Contributed Papers

1:30

1pPP1. A physiological model for comodulation masking release based on cross-frequency coincidence detection. Laurel H. Carney (Dept. of Bioengineering & Neurosci. and Inst. for Sensory Res., Syracuse Univ., 621 Skytop Rd., Syracuse, NY 13244, laurel_carney@isr.syr.edu)

Detection thresholds for tones in noise can be significantly improved when the masker is amplitude modulated, or when its various components are comodulated [e.g., Hall *et al.*, J. Acoust. Soc. Am. **76**, 50–56 (1984)]. Several stimulus-based signal-processing models have been proposed to explain this phenomenon, referred to as comodulation masking release (CMR); however, no physiologically based models have successfully described the psychophysical data. These models have tested hypotheses that CMR involves either cross-frequency processing strategies or within channel envelope-based processing strategies. Cross-frequency coincidence detection is a physiologically realistic, nonlinear processing mechanism that is consistent with suggestions that information from different frequency channels is combined to explain CMR. Furthermore, cross-frequency coincidence detectors are sensitive to envelope cues, thus this mechanism is also consistent with within-channel envelope-based models for CMR. The sensitivity of coincidence detectors to envelope cues is due to the nonlinear timing of low-frequency auditory-nerve (AN) fibers—changes in amplitude of the stimulus result in systematic changes in the relative times of phase-locked AN fibers tuned to different frequencies [Heinz *et al.*, J. Acoust. Soc. Am. **110**, 2065–2084 (2001)]. A cross-frequency coincidence-detection model can explain the psychophysical results for the basic CMR paradigm described in Hall *et al.* (1984). [Work supported by NIH-NIDCD.]

1:45

1pPP2. Cochlear filter nonlinearity between 500 and 8000 Hz in listeners with normal hearing. Enrique A. Lopez-Poveda (Centro Regional de Investigacin Biomédica, Facultad de Medicina, Universidad de Castilla, La Mancha, 02071 Albacete, Spain) and Christopher J. Plack (Univ. of Essex, Colchester CO4 3SQ, UK)

Human nonlinear cochlear responses were estimated in three normal-hearing listeners at characteristic frequencies (CF) of 0.5, 1.0, 2.0, 4.0, and 8.0 kHz, as a function of frequency and level. Cochlear responses were derived from fixed-probe-level isoresponse temporal masking curves (TMCs) at signal frequencies (F_s) equal to CF. Signal level was fixed at 14 dB SL. The level of the forward maskers was adjusted so as to find the masker level at threshold as a function of the time interval between the masker and the signal. Masker frequencies (F_m) of 0.5, 0.7, 0.9, 1.0, 1.1, and 1.6 F_s were used. Across CFs, results for $F_s \ll F_m$ show a single slope consistent with linear cochlear response. For $F_s = F_m$, TMCs show a two- or three-stage curve consistent with cochlear nonlinearity. A trend was observed in the compression slopes of derived cochlear IO functions as a function of CF, going from 0.8 dB/dB at 0.5 kHz, to 0.27 dB/dB at 8.0 kHz. Although shifts in the best frequency of the 0.5-, 1.0-, and 4.0-kHz cochlear filters were observed at moderate input levels, they were absent at higher levels. The data may be of use to tune models of the human basilar-membrane response.

2:00

1pPP3. Informational masking: Toward improved understanding. Nathaniel I. Durlach, Gerald Kidd, Jr., Christine R. Mason, Tanya L. Arbogast (Hearing Res. Ctr., Boston Univ., 635 Commonwealth Ave., Boston, MA 02215, durlach@mit.edu), H. Steven Colburn, and Barbara Shinn-Cunningham (Hearing Res. Ctr., Boston Univ., Boston, MA 02215)

This paper is concerned with simultaneous informational masking of tones by tone complexes. Past work distinguishes between energetic (peripheral) masking, which occurs when target and masker overlap in the peripheral frequency channels, and informational (central) masking, which can occur even when there is no such overlap but there is uncertainty in the frequency spectrum of the stimulus. We suggest that this distinction is (1) applicable to domains other than the frequency domain (e.g., the spatial domain) and (2) relative to the level of processing in the auditory system considered. We further suggest that this distinction can be rigorously defined for any domain and any level by comparing performance of the ideal detector operating on the inputs at that level to human psychophysical performance. We also argue that informational masking is strongly influenced by target-masker similarity and segregation difficulty, not only by stimulus uncertainty. Experiments designed to probe this issue are discussed in a companion paper (Mason *et al.*). Finally, ideas relevant to the large intersubject differences observed in informational masking (focused on the distinction between trying to hear the target and trying to not hear the masker) are used to make new predictions about these differences. [Work supported by NIH/NIDCD.]

2:15

1pPP4. Additivity of forward masking at low and high frequencies. Christopher J. Plack and Catherine G. O'Hanlon (Dept. of Psych., Univ. of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK, cplack@essex.ac.uk)

Recent physiological and psychophysical measurements suggest that the response of the cochlea is compressive at high frequencies, but much more linear at frequencies below around 500 Hz. In this experiment, the dependence of compression on frequency was investigated using the additivity of forward masking. Two equally effective forward maskers should produce 3 dB of additional masking when they are combined assuming linearity with respect to intensity. A combined effect greater than this (excess masking) indicates compression. The signal was a 10-ms pure tone presented 20 ms after a 200-ms narrow-band noise masker and/or immediately after a 20-ms narrow-band noise masker. The signal frequency was 250, 500, or 4000 Hz. The signal threshold produced by adding two equally effective maskers was measured. At all three frequencies, no excess masking was observed for a signal 10 dB above absolute threshold, indicating linear additivity (no compression). At signal levels 30 dB above absolute threshold, excess masking was observed at all three frequencies. The estimated compression exponent was 0.31 at 250 Hz, 0.35 at 500 Hz, and 0.17 at 4000 Hz. Although the results suggest less compression at low frequencies than high, the estimated compression at low frequencies was greater than that reported previously.

2:30

1pPP5. Intensity-difference limens for very brief Gaussian-shaped tones at 500 Hz and 6.5 kHz. Lance Nizami and Walt Jesteadt (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

The curve of intensity-difference limen (DL) versus level shows a midlevel hump for 6.5-kHz tones having 5-ms cosine-squared ramps and a 26-ms plateau [Carlyon and Moore, J. Acoust. Soc. Am. **76**, 1369–1376 (1984)]. DLs were presently obtained for that tone, as well as for 6.5-kHz tones having Gaussian envelopes, for five subjects. Pedestal levels covered 30- to 90-dB SPL and equivalent rectangular durations D were 0.314, 0.627, 1.25, 2.51, 10.03, and 20.1 ms, where D is the Gaussian-envelope duration divided by 3.19. The DLs for the tone of Carlyon and Moore ($D \sim 30$) show a midlevel hump smaller than found earlier. The hump grows noticeably with smaller D (i.e., the Gaussian-shaped tones), reaching 7 dB for $D=0.627$. The peak drops, however, for $D=0.314$, causing the curve of DL versus duration to peak for $D=0.627$ for tone levels of 40–60 dB SPL. DLs were also obtained from seven subjects for 500-Hz tones of 30–90 dB SPL for D's of 1.25, 2.51, 5.02, and 10.03 ms. A midlevel hump at D's of 1.25 and 2.51 gives way to a near-miss at D's of 5.02 and 10.03. A mid-duration peak is seen at $D=5.02$ for levels of 30–60 dB SPL.

2:45

1pPP6. Temporal interference in amplitude modulation perception. Jan Koopman (Dept. of Clinical and Exp. Audiol., Academic Medical Ctr., D2-223, P.O. Box 22660, 1100 DD, Amsterdam, The Netherlands), Tammo Houtgast (Free Univ. Medical Ctr., Amsterdam, The Netherlands), and Wouter Dreschler (Academic Medical Ctr., Amsterdam, The Netherlands)

We observed that, especially for hearing-impaired subjects, the perceived modulation strength of a signal is systematically reduced by the presence of a preceding modulated signal. This suggested that modulation detection interference (MDI) might not only apply in case of simultaneous signals, but also for nonsimultaneous signals. To further investigate this we performed an experiment with normal-hearing and hearing-impaired subjects, in which modulation detection was measured for a target signal (1 kHz carrier, 625 ms duration, 8 Hz amplitude modulation), preceded by an interfering signal (2 kHz carrier, 625 ms, 100% 8 Hz modulation). The onset difference between interfering signal and target signal was varied over a range of 0–1250 ms, thus from simultaneous (the traditional MDI condition) to clearly nonsimultaneous. Based on pilot experiments we hypothesized that normal-hearing subjects showed a rapid increase of sensitivity from simultaneous to nonsimultaneous presentation (essentially no MDI after effect), whereas for the hearing impaired the interference effect was observed up to long onset differences, including clearly nonsimultaneous presentation. The observed differences in modulation interference between normal-hearing and hearing-impaired subjects might explain the differences shown in speech intelligibility in fluctuating background noise. Results will be used to validate this hypothesis.

3:00–3:15 Break

3:15

1pPP7. Learning interference associated with training temporal-interval discrimination with two stimuli. Beverly A. Wright and Jeanette A. Ortiz (Dept. of Commun. Sci. and Disord. and Northwestern Univ. Inst. for Neurosci., Northwestern Univ., 2299 N. Campus Dr., Evanston, IL 60208-3550)

Determining the training regimes that do and do not yield improvements in auditory discrimination provides information about the mechanisms underlying the trained task, and about learning itself. In a previous report, training with one standard stimulus yielded improvements in temporal-interval discrimination that generalized to an untrained frequency at the trained interval [B. A. Wright *et al.*, J. Neurosci. **17**, 3956–3963 (1997)]. Here, the impact of training interval discrimination with two standard stimuli, as opposed to one, was investigated. Listeners were trained either with two standard intervals (100 and 350 ms) at one frequency (1 kHz; $n=8$), or with one standard interval (100 ms) at two frequencies (0.25 and 1 kHz; $n=8$). On average, listeners in both groups (1) learned less on each trained condition than listeners trained with only one stimulus ($n=10$), and (2) subsequently performed on the trained interval(s) at an untrained frequency (4 kHz) as though they had trained with only one of the stimuli—in one case, showing no improvement. These results indicate that, at some stage, temporal intervals are processed at frequency-specific sites, and that training one task with two stimuli sharing all but one feature can interfere with both learning and generalization. [Work supported by NIDCD.]

3:30

1pPP8. The auditory cross-section (AXS) test battery: A new way to study afferent/efferent relations linking body periphery (ear, voice, heart) with brainstem and cortex. Judith Lauter (Human Neurosci. Lab., Dept. of Human Services, Stephen F. Austin State Univ., Box 13019, SFA Station, Nacogdoches, TX 75962)

Several noninvasive methods are available for studying the neural bases of human sensory-motor function, but their cost is prohibitive for many researchers and clinicians. The auditory cross section (AXS) test battery utilizes relatively inexpensive methods, yet yields data that are at least equivalent, if not superior in some applications, to those generated by more expensive technologies. The acronym emphasizes access to axes—the battery makes it possible to assess dynamic physiological relations along all three body-brain axes: rostro-caudal (afferent/efferent), dorso-ventral, and right-left, on an individually-specific basis, extending from cortex to the periphery. For auditory studies, a three-level physiological ear-to-cortex profile is generated, utilizing (1) quantitative electroencephalography (qEEG); (2) the repeated evoked potentials version of the auditory brainstem response (REPs/ABR); and (3) otoacoustic emissions (OAEs). Battery procedures will be explained, and sample data presented illustrating correlated multilevel changes in ear, voice, heart, brainstem, and cortex in response to circadian rhythms, and challenges with substances such as antihistamines and Ritalin. Potential applications for the battery include studies of central auditory processing, reading problems, hyperactivity, neural bases of voice and speech motor control, neurocardiology, individually-specific responses to medications, and the physiological bases of tinnitus, hyperacusis, and related treatments.

3:45

1pPP9. Modulation masking in cochlear implant listeners: Effects of masker envelope. Monita Chatterjee (Dept. of Auditory Implants and Percept., House Ear Inst., 2100 W. Third St., Los Angeles, CA 90057)

We are interested in quantifying the effects of competing fluctuations on the detection of envelope modulation by cochlear implant listeners. In these experiments, the task is to detect modulation in the signal envelope in the presence of noisy or sinusoidal fluctuations in a masker envelope. Masker and signal pulse trains are both 500 Hz, 200 ms long and interleaved with each other. Noise is applied to the masker pulse train by scaling successive pulses by a random number drawn from a uniform distribution. The masker-signal tonotopic distance is varied. Results indicate that, for all tonotopic distances between the masker and signal tested, the fluctuating masker has a stronger masking effect on modulation detection than a steady-state masker with an amplitude fixed at the peak of the fluctuating envelope. Experiments with sinusoidally modulated maskers show significant effects of the phase and frequency of masker modulation relative to signal modulation. Overall, the results indicate significant interactions between concurrent, interleaved pulsatile stimuli in both tono-

topic and envelope domains. Therefore, measures of channel-interaction in cochlear implant listeners should incorporate the effects of the distance between the electrodes as well as the temporal envelope in each channel. [Work supported by NIDCD.]

4:00

1pPP10. When sound effects are better than the real thing. Laurie M. Heller and Lauren Wolf (Dept. of Psych., Box 1853, Brown Univ., Providence, RI 02912)

Sound effects technicians (''Foley Artists'') have long exploited the fact that two physically different events can produce perceptually similar sounds, such as squeezing a box of cornstarch to imitate footsteps in the snow. Although some sound effects succeed because they produce acoustic waveforms nearly identical to the sounds they are imitating (their targets), in other cases there are obvious acoustic differences between sound effects and their targets. Those differences may provide information about which acoustic features are essential, and which are extraneous, for auditory recognition of an event. To address this question, twelve pairs of sound effects and their associated target events were recorded. Listeners attempted to identify each sound as it was presented over headphones. Next, the listeners were informed of the target sound (e.g., footsteps in the snow) and were asked to rate the sound's realism. The sound effects were sometimes judged as more realistic than recordings of the target events. Acoustic differences between the waveforms of the most realistic sound effects and their targets were interpreted in terms of psychoacoustic principles.

4:15

1pPP11. Aging of the ear, exposed and not exposed to noise. Gerald Fleischer, Reinhard Müller, Guido Heppelmann, Thomas Bache (Dept. of Auditory Res., Univ. of Giessen, Aulweg 123, D-35392 Giessen, Germany), Xia Gao, Feng Wang, and Weiguo Huang (FMM Univ., Xian, Shaanxi 710 032, PROC)

Hearing in two special groups was compared, using pure-tone audiometry up to 16 kHz. One group consisted of commercial airline pilots in Germany, the other of people living in remote parts of the People's Republic of China, without electricity, cars or, other technical equipment. The age of the 502 pilots (99% male) ranged from 22 to 65 years, while that of the 644 Chinese (61% male) ranged from 6 to 82 years. In both groups participants with diseases of the ear and with acute acoustic trauma were excluded. Analysis showed that the threshold of hearing in young adults is identical in Germany and in China. It is evident that the effect of aging is reduced in pilots, as compared to people living without technical or industrial noise. Older pilots clearly hear better than people not subjected to technical noise, most prominently at higher frequencies. Together with other observations, it can be concluded that continuous wideband noise reduces the aging of the ear. Long-term exposure to such noise—within limits—appears to be good for the ear.

4:30

1pPP12. The effect of audiometric headphone position on the measurements of threshold levels. Jovani A. Steffani (West Univ. of Santa Catarina (UNOESC), Joacaba, SC, Brazil), Samir N. Y. Gerges (Federal Univ. of Santa Catarina (UFSC), Florianopolis, SC 88040-900, Brazil, samir@emc.ufsc.br), and Ana C. Fiorini (Univ. of Sao Camilo and PUC, Sao Paulo, Brazil)

The results of audiometric tests are sensitive to the position of the headphone on the ears. Up to 30-dB differences in all frequency bands can be obtained. In this paper measurements are reported for normal audiometric tests for the same audiometer equipment varying the listener and also for different audiometric equipment for the same listener. These measurements were also carried out using an artificial head with molded external ears to give more parameter control experimental results. All the measurements were carried out in an acoustic room with very low background noise. Nine positions were used for the measurements of SPL (one reference headphone position and four positions in the up, down, left, right and four other positions on the 45 deg), for each position three locations of the headphone were used at 5 mm, 10 mm and 15 mm away from the center), totaling 45 cases. Larger differences obtained at high frequencies. 72% of the measured cases give higher threshold levels and 28% gives lower levels than reference position. The results can give good contribution to the confidence limits and repeatability factor for audiometric testing.

Session 1eID

Please note: Payment of an additional registration fee is required to attend this lecture. See page A28

Interdisciplinary: Tutorial Lecture: Keep Your Ear on the Ball: Display of Targets in the Bat's Sonar Receiver

Joseph Pope, Chair
Pope Engineering Company, P.O. Box 590236, Newton, Massachusetts 02459-0002

Chair's Introduction—7:00

7:05

1eID1. Keep your ear on the ball: Display of targets in the bat's sonar receiver. James A. Simmons (Brown Univ., Providence, RI 02912, james_simmons@brown.edu)

This tutorial examines how the biological sonar systems of bats transform wideband acoustic information conveyed in echoes from their initial reception by the external ear and transduction by the inner ear through the cascade of processing stages in the auditory nervous system. This information finally is manifested as the content of the images the bat perceives. The intention of the tutorial is to give people having the diverse interests of ASA members—in acoustics and electronics engineering, signal processing, psychoacoustics, bioacoustics, and auditory physiology—a comprehensive view of how these images are formed and how all these areas of study prove vital for appreciating how unconventional is the organization of the bat's sonar receiver. The presentation will detail the various design problems associated with imaging of objects by high-resolution sonar, and how they appear to be solved by the bat's auditory system. Some critical functions are carried out by single structures, such as the middle ear, while other functions are so distributed that they are concealed from view at any one stage and can only be discerned by bringing together observations made across stages. The relevance of recent discoveries in cochlear physiology, time-frequency signal-processing, and neural display of information known to be retained in biosonar images will receive particular attention.

Session 2aAA

Architectural Acoustics: Adaptive Reuse of Vaudeville and Movie Palaces for Modern Music and Theater Performance

Carl P. Giegold, Chair
Kirkegaard Associates, 801 West Adams Street, Suite 800, Chicago, Illinois 60607

Chair's Introduction—10:30

Invited Papers

10:35

2aAA1. Acoustical components of the Orpheum Theatre renovation. David A. Conant (McKay Conant Brook, Inc., 5655 Lindero Cyn Rd., Ste. 325, Westlake Village, CA 91362, DConant@MCBinc.com)

The before and after acoustically-important listening conditions and measurements are described for a $14.2 million restoration and renovation of this 1,400-seat grand 1929 Movie Palace—at one time the tallest building in Phoenix. Great care was taken to restore all the acoustically good parts of the original design and to subtly modify the acoustically-troublesome parts (including severely focusing dome and sidewalls) so they looked the same but performed properly. A beautiful playhouse was achieved in 1997 with fine, fixed acoustics and conventional audio reinforcement, a surrounding Italian village, plenty of gilt, moving clouds and sunsets overhead. Today, the venue successfully hosts shows from performance art to ballet to jazz. It was not intended (acoustically) to serve classical music but does so on occasion.

10:55

2aAA2. A case study: Transforming a vaudeville theatre into a 21st century multi-purpose hall. Richard H. Talaske, Ian B. Hoffman, and Jonathan P. Laney (The Talaske Group, Inc., 105 N. Oak Park Ave., Oak Park, IL, rick@talaske.com)

Originally constructed in 1927 as an atmospheric, vaudeville theatre, the Coronado enjoyed 50 years as host to a true variety of events. In the 1970s, functional and space limitations became a catalyst for touring events to avoid the Coronado, and the theatre fell into disrepair. In the mid-1990s, a renovation was planned (completed January 2001) to restore the glory of the original theatre and upgrade it to be a viable multi-purpose hall for the 21st century. By annexing two neighboring buildings, many of the space limitations could be overcome in both the front- and back-of-house. The scope included a significantly enlarged stage and stagehouse, performer support spaces, expanded lobby spaces, quiet ventilation systems, and a new music rehearsal room. Acoustically, a primary goal was to upgrade the hall for both reinforced music/theatre performances and unreinforced, orchestral performances. However, historical regulations prevented modification to the room shape and configuration. In order to make the room suitable for unreinforced music, two primary acoustic features were instituted—a semi-custom concert enclosure for the stage and an acoustic enhancement system for the audience chamber. For theatrical and popular music events, a new performance audio system was designed and integrated into the hall.

11:15

2aAA3. Vaudeville-era renovations from the architect's point of view. Albert L. Filoni (MacLachlan, Cornelius & Filoni, Inc., 200 The Bank Tower, 307 4th Ave., Pittsburgh, PA 15222, afiloni@mcfarchitects.com)

Interaction between the architect and acoustics consultant on renovation and restoration projects often must reconcile recommendations critical to acoustics goals with uncooperative existing architecture or unwilling landmarks jurisdictions. Through case studies including the renovation of the Pittsburgh Symphony Orchestra's Heinz Hall, the author will discuss the relationship between the architect and acoustics consultant in resolving the design issues inherent in vaudeville-era buildings.

Contributed Papers

11:35

2aAA4. The renovation of Madison's Oscar Mayer Theater. Joseph W. A. Myers and Paul T. Calamia (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

Originally opened in 1928 as the Capitol Theatre, the Oscar Mayer Theatre in Madison, WI underwent substantial renovation in 1980 to support its current performance program. The theatre is now home to the Madison Symphony and the Madison Opera, and is used for ballet, touring shows, and popular concerts as well. As part of the ongoing Overture Project which will transform the theatre's home, the Madison Civic Center, into the Overture Center, the Oscar Mayer Theatre will be further improved to support its new role. In this paper we will discuss the reasoning behind the upcoming renovation within the context of the Overture Project, we will describe the pending architectural modifications to the theatre, and we will discuss the intended changes in acoustics. Computer modeling results will also be presented for the existing and renovated conditions.

11:50

2aAA5. There's nothing Majestic about burlesque. Christopher Jaffe and Robin Glosemeyer (Jaffe Holden Acoust., 114A Washington St., Norwalk, CT 06854)

Turn-of-the-century vaudeville and movie palaces initially seem to lend themselves to use as concert halls or multipurpose theatres. Basic features, including existing structure, appropriate location, and desirable seat count, already exist. But, upon further investigation, these palaces present both acoustical and theatrical limitations. Specifically, shallow stage depth, undersized stage wings, small orchestra pit, deep balcony overhangs, and low ceiling height in stage house and/or audience chamber are common deficiencies which must be addressed in successful conversions. Despite these challenges, converting an existing building is often a viable option for many communities. This paper will investigate several representative palaces which were modified to accommodate modern symphonic and theatrical productions between 1959 and the present.

TUESDAY MORNING, 4 JUNE 2002 LE BATEAU ROOM, 8:00 TO 11:35 A.M.

Session 2aAB

Animal Bioacoustics: Session on Marine Mammal Bioacoustics in Honor of William E. Evans I

Jeanette A. Thomas, Cochair
Laboratory of Sensory Biology, Western Illinois University, 3561 60th Street, Moline, Illinois 61265

Ann E. Bowles, Cochair
Hubbs-Sea World Research Institute, 2595 Ingraham Street, San Diego, California 92109

Chair's Introduction—8:00

Invited Papers

8:15

2aAB1. Perception time and movement time in dolphin pulsing and whistling. Sam Ridgway and Donald Carder (SPAWAR Systems Ctr. San Diego, Div. D235, 53560 Hull St., San Diego, CA 92152-5001)

Auditory/vocal response time was separated into perception time (PT) and movement time (MT) in trials with bottlenose dolphins (*Tursiops truncatus*)—two males and one female. Pressure catheters accepted into the nasal cavity by each dolphin recorded the pressure increase that preceded sound production. Time from acoustic stimulus onset to onset of pressure rise was recorded as PT (range 57 to 314 ms) and pressure rise onset to dolphin sound onset was recorded as MT (range 63 to 363 ms). Blindfolded dolphins trained to report a target by whistling often responded before completion of their 200- to 800-ms echolocation click trains. Detection of the target, indicated by whistling, before termination of the animal's own click train, suggests that dolphins do not voluntarily respond to each successive click but rather set a rhythm such that each click is emitted about 20 ms after the target echo arrives.

8:40

2aAB2. Interactions between commercial fishing and walleye pollock aggregations. Sarah Stienessen, Chris D. Wilson, and Anne B. Hallowed (Alaska Fisheries Sci. Ctr., Natl. Ocean. and Atmos. Admin., 7600 Sand Point Way, NE, Seattle, WA 98115, sarah.stienessen@noaa.gov)

Scientists with the Alaska Fisheries Science Center are conducting a multiyear field experiment off the eastern side of Kodiak Island in the Gulf of Alaska to determine whether commercial fishing activities significantly affect the distribution and abundance of walleye pollock (*Theragra chalcogramma*), an important prey species of endangered Steller sea lions (*Eumetopias jubatus*). In support of this activity, spatio–temporal patterns were described for pollock aggregations. Acoustic–trawl surveys were conducted in two adjacent submarine troughs in August 2001. One trough served as a control site where fishing was prohibited and the other as a treatment site where fishing was allowed. Software, which included patch recognition algorithms, was used to extract acoustic data and generate patch size and shape-related variables to analyze fish aggregations. Important patch related descriptors included skewness, kurtosis, length, height, and density. Estimates of patch fractal dimensions, which relate school perimeter to school area, were less for juvenile than for adult aggregations, indicating a more complex school shape for adults. Comparisons of other patch descriptors were made between troughs and in the presence and absence of the fishery to determine whether trends in pollock aggregation dynamics were a result of the fishery or of naturally occurring events.

9:05

2aAB3. Age-related hearing loss in sea lions and their scientists. Ronald J. Schusterman, Brandon Southall, David Kastak, and Colleen Reichmuth Kastak (Long Marine Lab., Univ. of California at Santa Cruz, 100 Shaffer Rd., Santa Cruz, CA 95060)

Interest in the hearing capabilities of California sea lions (*Zalophus californianus*) was first stimulated by the echolocation hypothesis and more recently by rising concern about coastal noise pollution. During a series of audiometric tests, we measured the absolute hearing sensitivity of two sea lions and two of their human investigators. Aerial hearing curves for each subject were obtained with a go/no-go procedure and standard psychophysics. Additionally, underwater hearing curves were obtained for the sea lions using

the same procedures. Underwater, the older sea lion (22–25 years of age) showed hearing losses relative to the younger sea lion (13–16 years) that ranged from 10 dB at lower frequencies to 50 dB near the upper frequency limit. The older sea lions' hearing losses in air were consistent with those measured underwater. The older human (69 years) tested also showed losses relative to the younger human (22 years). These differences ranged from 15 dB at lower frequencies up to 35 dB at the highest frequency tested. The results obtained in this study document age-related hearing losses in sea lions and humans. The findings are consistent with data on presbycusis in other mammalian species, showing that maximum hearing loss occurs at the highest frequencies.

9:30

2aAB4. Interrelationships between intranarial pressure and biosonar clicks in the bottlenose dolphin (*Tursiops truncatus*). Wesley R. Elsberry (Marine Acoust. Lab., Texas A&M Univ., Galveston, TX), Ted W. Cranford (San Diego State Univ., San Diego, CA), Sam H. Ridgway (SPAWAR Syst. Ctr., San Diego, CA), Donald A. Carder, William G. VanBonn (US Navy Marine Mammal Prog., San Diego, CA), Diane J. Blackwood (Marine Acoust. Lab., Texas A&M Univ., Galveston, TX), Jennifer A. Carr (US Navy Marine Mammal Prog., San Diego, CA), and William E. Evans (Texas A&M Univ., Galveston, TX)

Three Atlantic bottlenose dolphins (*Tursiops truncatus*) were given a target recognition biosonar task. During their performance of the task, both acoustic data in the far field and pressure within the bony nasal passages were digitally recorded (Elsberry *et al.*, 1999). Analysis of over 15 000 biosonar clicks provided new insights into odontocete biosonar sound production and is consistent with acoustic and pressure data taken from white whales (*Delphinapterus leucas*) during biosonar (Ridgway and Carder, 1988). Our work provides the first evidence for a minimum intranarial pressure during biosonar click production for any odontocete (11.8 ± 0.5 kPascals over basal pressure). All three subjects exhibited nearly the same minimum tranarial pressure difference during biosonar click production. Clicks produced at or near this minimum intranarial pressure exhibited a wide range of acoustic power values. The acoustic power of a biosonar click was not highly correlated with intranarial pressure ($R^2 = 0.116$). The radiated acoustic energy in biosonar clicks ranged from 1 to 1370 microJoules. Estimates of mechanical work during pressurization events were produced using a piston/cylinder model and intranarial volume data from prepared specimens and computed tomography scans. Mechanical work during pressurization events ranged from 2.74 to 23.0 Joules, with an average of 10.3 Joules.

9:55

2aAB5. Past and present research on gray whale vocalizations. Francisco Ollervides (RARE Ctr. for Tropical Conservation, 1840 Wilson Blvd., Ste. 402, Arlington, VA 22201-3000, follervides@rarecenter.org)

Eberhadt and Evans first recorded vocalizations of gray whales in 1967 in Laguna Ojo de Liebre, Mexico. Three and a half decades of bioacoustic research on gray whales has followed this groundbreaking work. Gray whales appear more vocal while at the southern breeding lagoons off Baja California and are least vocal at the northern feeding areas of the Bering, Chukchi, and Beaufort Seas. Ten different vocalizations have been identified on the breeding grounds, while only six different vocalizations have been recorded on the feeding grounds. Evans also conducted the first recordings of vocalizations of gray whales in captivity with his work on J.J. II in San Diego in 1974. Since then, the database of captive gray whale vocalizations has expanded with research on GiGi in 1997. From this review, an acoustic repertoire of at least 12 different types of vocalizations is suggested. Nevertheless, not all gray whale calls fall into obvious or distinct categories; thus, correlating specific behaviors with each vocalization type merits further study.

10:20–10:30 Break

10:30

2aAB6. Do manatees talk during sex? Caryn Self-Sullivan (Texas A&M Univ., College Station, TX 77843-2258), Tamra Gilbertson (Mote Marine Lab., Sarasota, FL 34236), and William E. Evans (Texas A&M Univ., Galveston, TX 77553-1675)

On January 13, 1999, manatee vocalizations were recorded during a mating herd event in the Orange River, Florida. Although copulation could not be observed, multiple males were observed with exposed penises. During one 25 min sample (1300–1325 h), over 400 manatee signals were recorded. In March 2000, each signal was captured and digitized from the analog tape using a Marantz PMD 501, Ashly equalizer (gain=0, filter=0), MAC 8100, and Canary 1.2.1. In general, signals were 100–200 ms in length, highly harmonic (up to 8 harmonics ranging from 1 to 16 kHz), with little or no frequency modulation. Intervals between signals ranged from less than 1 s to 14 s (mean = 3 s), indicating that manatees do indeed talk (a lot) during sex. Noise from two passing boats was also recorded during the sample period. One abnormally low-frequency signal (0.4 kHz) was recorded during one boat pass. This apparent manatee vocalization could be seen and heard below the boat noise frequency band.

10:55

2aAB7. Echolocation signals of foraging killer whales (*Orcinus orca*). Whitlow W. L. Au (Hawaii Inst. of Marine Biol., Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734), John K. B. Ford (Pacific Biological Station, DFO Canada, Nanaimo, BC V6B 3X8, Canada), and Kelly A. Allman (The Marine Mammal Ctr., Sausalito, CA 94965)

Fish eating resident killer whales that frequent the coastal waters of Vancouver Island, Canada have a strong preference for chinook salmon. The whales in Johnston Strait often forage along the steep cliffs that extend into the water, echolocating their prey. Echolocation signals were measured with a four hydrophone symmetrical star array and the signals were simultaneous digitized at a sample rate of 500 kHz using a lunch-box PC. A portable VCR recorded the images from an underwater camera located close to the array center. Only signals emanated from close to the beam axis (1185 total) were chosen for a detailed analysis. Killer whales project very broad band echolocation signals (Q 1.3 to 1.5) that tend to have a bimodal frequency structure. Ninety seven percent of the

signals had center frequencies between 45 and 80 kHz with a band-width between 35 and 50 kHz. The peak-to-peak source level of the echolocation signal decreased as a function of the one way transmission loss to the array. Source levels varied between 200 and 225 dB re 1 μPa. Using a model of target strength for chinook salmons, the echo levels from the echolocation signals are estimated for different ranges between whale and salmon.

Contributed Paper

11:20

2aAB8. Evoked potential application to study of echolocation in cetaceans. Alexander Ya. Supin (Inst. of Ecology and Evolution, Russian Acad. of Sci., 33 Leninsky Prosp., 119071 Moscow, Russia, a.supin@g23.relcom.ru), Paul E. Nactigall, Jeffrey Pawloski, and Whitlow W. L. Au (Hawaii Inst. of Marine Biol., Univ. of Hawaii, Kailua, HI)

The evoked-potential (EP) method is effective in studies of hearing capabilities of cetaceans. However, until now EP studies in cetaceans were performed only in conditions of passive hearing by recording EP to external stimuli. Can this method be applied to study active echolocation in odontocetes? To answer this question, auditory brainstem evoked responses (ABR) were recorded in a false killer whale while the animal echolocated a target within an experiment in which the animal reported the target present or absent. The ABR collection was triggered by echolocation clicks. In these conditions, the recorded ABR pattern contained a duplicate set of waves. A comparison of ABR wave delays recorded during echolocation with those recorded during regular external stimulation has shown that the first set of waves is a response to the emitted click whereas the second one is a response to the echo. Both responses, to the emitted click and to the echo, were of comparable amplitude in spite of the intensity difference of these two sounds of more than 40 dB near the animal's head. This finding indicates some mechanisms releasing responses to echoes from masking by loud emitted clicks. The evoked-potential method may be productive to investigate these mechanisms.

TUESDAY MORNING, 4 JUNE 2002 GRAND BALLROOM 3, 8:55 TO 11:35 A.M.

Session 2aAO

Acoustical Oceanography and Underwater Acoustics: Acoustical Studies of Bubbles in the Ocean

R. Lee Culver, Chair

Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

Chair's Introduction—8:55

Invited Papers

9:00

2aAO1. Use of bubble cloud attenuation measurements to determine the bubble size distribution and comparison with backscatter measurements. Hugh Dumbrell (Offshore & Acoust. Dept., QinetiQ Ltd., Bincleaves Technol. Park, Newtons Rd., Weymouth, Dorset DT4 8UR, UK, hadumbrell@qinetiq.com)

This paper presents the method of using bubble cloud attenuation measurements to determine the bubble size distribution, taking into account the effect of time-dependent bubble scatter and absorption. It then shows that the results of attenuation measurements due to bubble clouds (generated by a ship's wake) agree with backscatter measurements made at a much higher frequency.

9:20

2aAO2. Acoustic measurements of the number of naturally produced bubbles resonating at 50 kHz. Bernd Nuetzel and Karl-Heinz Herwig (FWG, Klausdorfer Weg 2-24, 24148 Kiel, Germany)

When the wind is strong enough to form whitecaps, bubbles caused by breaking waves penetrate into the water column and are transported to depths of more than 10 meters by Langmuir circulation, turbulence, and other mechanisms. In the near-surface layer, bubbles cause acoustic scattering and attenuation. Especially in shallow water, where the sound is reflected and scattered at the boundaries (sea bottom and surface), bubbles significantly influence sound propagation in the sea. The results from acoustic near-surface experiments conducted in the North Sea over several years are presented. Data sets of acoustic attenuation were measured inside the bubble clouds for a propagation range of 1 m. Also, a narrow-beam transducer was utilized to measure the acoustic scattering from the near-surface bubbles. Both investigations were done simultaneously for a large variety of environmental conditions. The data sets have been combined to determine the number of resonating bubbles as a function of depth and wind speed for a selected frequency of 50 kHz. A numerical expression is introduced for this special set of parameters. The comparison with results from other investigators using different techniques shows good agreement.

9:40

2aAO3. Laboratory studies on scattering from bubbles located close to an air–water interface. George Kapodistrias (Philips Medical Systems, Bothell, WA 98021, george.kapodistrias@philips.com) and Peter H. Dahl (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

Results of laboratory experiments addressing the problem of scattering from bubbles located near an air–water interface are discussed. Well-controlled measurements of scattering from bubbles placed on a fine thread (itself nearly transparent to sound) were made. Depending on the experiment, bubble radii ranged from 500–1000 μm, while insonifying frequencies ranged from 110–130 kHz. The experiments quantified multiple scattering effects, by obtaining scattering data from two bubbles symmetrically arranged at a distance, d, from the acoustic beam axis, and the influence of both a flat and a roughened air–water interface, by measuring backscattering from a bubble located a distance, d, from the interface. Experiments were performed in the kd range 0.2–21, where k is the acoustic wave number. Modeling was achieved by utilizing the multiple scattering series and ray synthesis of the four scattering paths, created by the proximity of the bubbles to the interface, while addressing both bistatic and monostatic conditions. Experimental results are in good agreement with simulations, and they highlight three important features of the problem of acoustic observations from subsurface bubbles: (1) multiple propagation paths, (2) multiple scattering, and (3) rough surface effects. The relation of this work to interpretive models of field data is also discussed.

10:00

2aAO4. The acoustics of bubbles and bubbly assemblages: a suite of laboratory studies. Ronald A. Roy, Preston S. Wilson, Ryan D. McCormick, Eun-Joo Park (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215, ronroy@bu.edu), and William M. Carey (Boston Univ., Boston, MA 02215)

The *in situ* study of oceanic bubble acoustics is a challenging endeavor. One must impose a variety of *a priori* assumptions and constraints in order to extract information that unambiguously sheds light on the underlying physical mechanisms that govern the transmission and scattering of sound in oceanic bubble assemblages. Conversely, laboratory studies afford an added measure of physical control and observational access that permits one to focus on certain facets of bubble activity with accuracy and precision. We provide an overview of several laboratory-based bubble acoustics studies under way at Boston University. The scope of the work spans the spectrum from the quantitative measurement of acoustic dispersion through, and scattering from, high void fraction bubbly regions, to detailed studies of the linear dynamics of individual bubbles. [Work supported by ONR.]

10:20–10:35 Break

Contributed Papers

10:35

2aAO5. Measuring bubble distribution with multibeam sonar. R. Lee Culver and David Bradley (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804, rlc5@psu.edu)

Bubbles in water are very efficient scatterers of sound. Depending upon the ratio of acoustic wavelength to bubble radius, the scattering mechanism can be Rayleigh, resonant, or geometric. The strength of the scattered signal depends strongly on this ratio. The strength of the scattered signal can be used to estimate the number of bubbles per unit volume, or bubble density, if the size of the bubbles or the size distribution of the bubbles is known. A number of researchers have employed narrow beam sonar to measure the density of bubbles in water. We have used a 250 kHz sonar with 19 narrow beams, integrated into an autonomous vehicle (AUV), to image the wakes of two large surface ships from the underside. An earlier talk focused on the structure of the turbulent wake revealed by the measurements [J. Acoust. Soc. Am. **110**, 2642 (2001)]. In this talk we discuss parameters and assumptions which strongly affect the bubble density estimate. These include: values for the transmit and receive beamwidth; the shape of the bubble size distribution spectrum, including possible variation with depth; strong boundary reflections and high sidelobe levels; and the possible influence of multiple scattering. [Work sponsored by ONR Code 321OA.]

10:50

2aAO6. Multibeam echo-sounding measurement of the microbubble field in a ship's wake. Timothy C. Gallaudet (Marine Physical Lab., Scripps Inst. of Oceanogr., 9500 Gilman Dr., La Jolla, CA 92093-0205) and Christian P. de Moustier (Univ. of New Hampshire, Durham, NH 03824)

Acoustic data collected with the U.S. Navy's 68 kHz Toroidal Volume Search Sonar (TVSS) system were used to measure the microbubble field generated by the wake of the towing vessel at a constant distance of 18 ship lengths astern. Swath bathymetry beamforming and echo detection techniques were adapted to map the sea surface and the spatial distribution of bubble fields observable in the ocean volume with the TVSS. The sonar was towed at 78 m depth yielding a sea surface measurement swath roughly 550 m wide thus allowing observations of wake bubble fields across three parallel ship tracks spaced 200 m apart and over 22 min after generation. Measurements in microbubble fields about 3 min after generation by the ship's wake yielded an average vertical attenuation of 2.1 dB/m and a maximum acoustic volume scattering strength of -17.1 dB. [Work supported by the U.S. Naval Research Laboratory.]

11:05

2aAO7. Characteristics of open ocean bubble plumes from a 3-month inverted echo-sounder record. Mark V. Trevorrow (Defence Res. Establishment Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada)

A self-contained 200-kHz inverted echo-sounder was deployed at Ocean Station Papa (NE Pacific, 1400 km west of Vancouver Is.) from 1 March to 20 May, 1996. The instrument operated continuously, recording calibrated volume scattering from near-surface bubbles with 3-s and 30-cm resolution. Average surface wind and significant wave height during this period were 7.5 m/s and 2.8 m, respectively. The data show the frequent occurrence of bubbles organized into vertical, plume-like structures, presumably drawn downward within the convergence zones of Langmuir circulation. Significant bubble plume activity (defined as average plume depth >3 m) was observed during 61% of the operating period. Average bubble plume penetrations of up to 15 m were observed, with maximum penetrations up to 25 m. Within the plumes, the backscatter cross section exhibited an exponential decay with depth, with e-folding scale in the range 0.5 to 2.5 m. Some implications of this plume-like structure for high-frequency sonar performance will be discussed.

11:20

2aAO8. Measurements of the time-dependent attenuation in a nonstationary bubble distribution. Preston S. Wilson, Ronald A. Roy, and William M. Carey (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02115)

In the ocean, natural and artificial processes generate clouds of bubbles which scatter and attenuate sound, and the structure of these clouds (i.e., the space-dependent bubble size distribution and void fraction) evolves over time. Measurements have shown that at the individual bubble resonance frequency (IBRF), sound propagation in this medium is highly attenuated and dispersive. Theory to explain this behavior exists in the literature and is experimentally verified away from resonance. However, due to excessive attenuation near resonance and the high degree of sensitivity to bubble population parameters, little adequate experimental data exists for comparison and theory remains largely unverified. Using an impedance tube, the time-dependent attenuation has been measured in a nonstationary laboratory bubbly fluid at IBRF, along with measurements of the accompanying time-dependent bubble population parameters. The relative variability of the measured maximum attenuation and the frequency at which it occurred agreed closely with existing theory. [Work supported by ONR.]

TUESDAY MORNING, 4 JUNE 2002 BRIGADE ROOM, 8:00 TO 10:05 A.M.

Session 2aEA

Engineering Acoustics and Architectural Acoustics: Electroacoustic Systems for 3-D Audio

Gary W. Elko, Chair
MH Acoustics, 26 Blackburn Place, Summit, New Jersey 07901

Chair's Introduction—8:00

Invited Papers

8:05

2aEA1. Measurement of 3-D directional characteristics of reverberant sound fields using a spherical microphone array. Bradford N. Gover (Inst. for Microstructural Sci., Natl. Res. Council, Ottawa, ON K1A 0R6, Canada, bradg@audiolab.uwaterloo.ca)

A new measurement system has been used to analyze directional characteristics of reverberant sound fields in several small- and medium-sized rooms. The system employs a pair of 32-element spherical microphone arrays to obtain the room impulse response in each of 60 steering directions. The two different-sized arrays are used to cover the frequency range from 300–3300 Hz with an angular (directional) resolution of about 28 degrees. An analysis of these directional room impulse responses enables the variation of arriving sound energy with (3-D) direction and time to be studied. Specifically, the isotropy of the sound field at the receiving position can be assessed at any time instant, or over any time range desired. Directions from which significant or insignificant proportions of sound energy arrive are identifiable. Furthermore, the time and direction of incidence of the direct sound and discrete early reflections can be identified. This can be of use in localizing sources, finding reflections, and in decomposing the sound field into arriving components.

8:35

2aEA2. A spherical microphone array for spatial sound recording. Jens Meyer and Gary W. Elko (mh acoustics, jm@mhacoustics.com, gwe@mhacoustics.com)

The progression of audio from monophonic to the present day 5-channel playback is being driven by the desire to improve the immersion of the listener into the acoustic scene. In the limit, the goal is the reconstruction of the original sound field. This talk describes the decomposition of the sound field into orthogonal components, the so-called spherical harmonics and is directly related to the method of Ambisonics. These components contain all required information to allow a reconstruction of the original sound field. This approach is scalable to any number of loudspeakers and is also backwards compatible to surround sound, stereo and mono playback. One problem is the recording of the orthogonal components. So far only solutions exist that allow the recording of spherical harmonics up to first order. This limits the spatial resolution. This presentation introduces a new microphone that overcomes this limitation. It consists of pressure sensors that are equally distributed on the surface of a rigid sphere. The number of sensors depends on the highest order spherical harmonic to be recorded. A minimum of $(n+1)^2$ sensors is required to record harmonics up to nth order. The sensor signals are then processed to give the desired spherical harmonic outputs.

9:05

2aEA3. Control transducer locations for virtual acoustic imaging using binaural principle. Takashi Takeuchi (Bldg. Environment Group, Bldg. Eng. Dept., Kajima Tech. Res. Inst., 2-19-1 Tobitakyu, Chofu-shi Tokyo 182-0036, Japan) and Philip A. Nelson (Univ. of Southampton, England)

When binaural sound signals are presented with loudspeakers, the systems inversion involved gives rise to a number of problems such as loss of dynamic range, deterioration of control performance by small errors, and room reflections. A method of overcoming these fundamental problems is proposed. A conceptual monopole transducer is introduced whose azimuthal position varies continuously as frequency varies. This gives a minimum processing requirement of the binaural signals for the control to be achieved and all

the above problems either disappear or are minimized. Among a number of practical solutions to realize such optimally distributed transducers is a discretization that enables the use of conventional transducer units. As for the elevation location of the control transducers, the analysis of the spectral cues and dynamic cues is performed. The frequency response of the plant that relates transducer outputs to ear-pressure signals suggests that control transducer positions will be promising at positions in the frontal plane above the listener's head. The analysis of the dynamic cues induced by unwanted head rotation also strongly supports the use of transducer locations in the frontal plane. Subjective experiments are performed and the methods described here are proved to be advantageous.

9:35

2aEA4. Audio recording and reproduction in CARROUSO: Getting closer to perfection? Heinz Teutsch, Sascha Spors, Herbert Buchner, Rudolf Rabenstein, and Walter Kellermann (Chair of Multimedia Commun. and Signal Processing, Univ. of Erlangen-Nuremberg, Cauerstr. 7, 91058 Erlangen, Germany)

State-of-the-art systems for spatial audio reproduction utilize two to six discrete playback channels. A problem inherent to these systems is the relatively small area where the listener is able to experience a true 3-D sound sensation. This so-called ''sweet spot'' can be significantly enlarged by using loudspeaker arrays in combination with wave field synthesis (WFS) technology, initially developed at Delft University. By following this approach, actual sonic spaces can be reproduced in their entirety and not only discrete multichannel representations thereof. While loudspeaker arrays can be used to reproduce sound fields, microphone arrays can be used for sound field capture and analysis. Having high-quality audio reproduction in mind, microphone array designs are presented that need to fulfill stricter requirements than what has been traditionally considered for microphone array applications. Information on acoustic source position is essential for WFS-based rendering techniques. As will be shown, joint audio–video object tracking proves to be efficient for this task. Moreover, full-duplex applications based on WFS technology, like high-quality teleconferencing or remote music teaching, call for sophisticated multichannel acoustic echo cancellation algorithms. The European project ''CARROUSO'' aims at developing, integrating, and building a real-time system that embraces all previously described technologies in an MPEG-4 context.

TUESDAY MORNING, 4 JUNE 2002 GRAND BALLROOM 1, 10:00 A.M. TO 12:00 NOON

Session 2aED

Education in Acoustics: Hands-on Demonstrations in Acoustics for High School Students

Uwe J. Hansen, Chair

Department of Physics, Indiana State University, Terre Haute, Indiana 47809

Approximately 20 experiments designed to introduce various acoustics principles will be set up in the room. After a brief introduction of each demonstration, they will be available for experimentation by visiting high school students. Space permitting, conference participants are also invited to view the demonstrations and perform hands-on experiments, provided such activity does not interfere with the purpose of the session, i.e. to introduce acoustics experiments to high school students. Experiments performed by the students will range in sophistication from simple wave studies to mapping of normal modes in two-dimensional structures.

TUESDAY MORNING, 4 JUNE 2002 RIVERS ROOM, 8:25 TO 11:30 A.M.

Session 2aMU

Musical Acoustics: Interactive Computer Music Systems

Roger Dannenberg, Chair

School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Chair's Introduction—8:25

Invited Papers

8:30

2aMU1. The structural implications of interactive creativity. Joel Chadabe (Electron. Music Foundation, 116 N. Lake Ave., Albany, NY 12206, chadabe@aya.yale.edu)

The functioning of any particular electronic musical instrument can be placed somewhere along a line that extends from deterministic to indeterministic. Although deterministic instruments may offer more powerful controls than traditional instruments, they typically put a performer in the traditional situation of making a gesture and expecting a predictable effect. Indeterministic instruments, on the other hand, put a performer in an interactive role of improvising relative to an unpredictable output. The unique

advantage of such interactive instruments is that they foster ''interactive creativity.'' The design of a traditional instrument is fundamentally different from the design of an interactive instrument. A traditional instrument is structured as a single cause and effect, articulated as a synchronous linear path through a hierarchy of controls from a performer operating an input device to the multiple variables of a sound generator. An interactive instrument, on the other hand, is structured as a network of many causes and effects at various levels of importance, with a performer's input as only one of the causes of the instrument's output in sound. The author will present several historical examples of interactive electronic musical instruments and offer some speculations on the future.

9:00

2aMU2. Live interactive computer music performance practice. David Wessel (Ctr. for New Music and Audio Technologies (CNMAT), Dept. of Music, Univ. of California, Berkeley, Berkeley, CA 94720, wessel@cnmat.berkeley.edu)

A live-performance musical instrument can be assembled around current lap-top computer technology. One adds a controller such as a keyboard or other gestural input device, a sound diffusion system, some form of connectivity processor(s) providing for audio I/O and gestural controller input, and reactive real-time native signal processing software. A system consisting of a hand gesture controller; software for gesture analysis and mapping, machine listening, composition, and sound synthesis; and a controllable radiation pattern loudspeaker are described. Interactivity begins in the set up wherein the speaker–room combination is tuned with an LMS procedure. This system was designed for improvisation. It is argued that software suitable for carrying out an improvised musical dialog with another performer poses special challenges. The processes underlying the generation of musical material must be very adaptable, capable of rapid changes in musical direction. Machine listening techniques are used to help the performer adapt to new contexts. Machine learning can play an important role in the development of such systems. In the end, as with any musical instrument, human skill is essential. Practice is required not only for the development of musically appropriate human motor programs but for the adaptation of the computer-based instrument as well.

9:30

2aMU3. The IMUTUS interactive music tuition system. George Tambouratzis, Stelios Bakamidis, Ioannis Dologlou, George Carayannis, and Markos Dendrinos (Inst. for Lang. and Speech Processing, 6 Artemidos Str. & Epidavrou, 15125 Paradissos Amaroussiou, Greece)

This presentation focuses on the IMUTUS project, which concerns the creation of an innovative method for training users on traditional musical instruments with no MIDI (Musical Instrument Digital Interface) output. The entities collaborating in IMUTUS are ILSP (coordinator), EXODUS, SYSTEMA, DSI, SMF, GRAME, and KTH. The IMUTUS effectiveness is enhanced via an advanced user interface incorporating multimedia techniques. Internet plays a pivotal role during training, the student receiving guidance over the net from a specially created teacher group. Interactiveness is emphasized via automatic-scoring tools, which provide fast yet accurate feedback to the user, while virtual reality methods assist the student in perfecting his technique. IMUTUS incorporates specialized recognition technology for the transformation of acoustic signals and music scores to MIDI format and incorporation in the training process. This process is enhanced by periodically enriching the score database, while customization to each user's requirements is supported. This work is partially supported by European Community under the Information Society Technology (IST) RTD programme. The authors are solely responsible for the content of this communication. It does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of data appearing therein.

10:00

2aMU4. Machine musicianship. Robert Rowe (New York Univ., 35 W. Fourth St., Rm. 777, New York, NY 10012, robert.rowe@nyu.edu)

The training of musicians begins by teaching basic musical concepts, a collection of knowledge commonly known as musicianship. Computer programs designed to implement musical skills (e.g., to make sense of what they hear, perform music expressively, or compose convincing pieces) can similarly benefit from access to a fundamental level of musicianship. Recent research in music cognition, artificial intelligence, and music theory has produced a repertoire of techniques that can make the behavior of computer programs more musical. Many of these were presented in a recently published book/CD-ROM entitled *Machine Musicianship*. For use in interactive music systems, we are interested in those which are fast enough to run in real time and that need only make reference to the material as it appears in sequence. This talk will review several applications that are able to identify the tonal center of musical material during performance. Beyond this specific task, the design of real-time algorithmic listening through the concurrent operation of several connected analyzers is examined. The presentation includes discussion of a library of C++ objects that can be combined to perform interactive listening and a demonstration of their capability.

10:30

2aMU5. Making the computer "listen" to music. Christopher S. Raphael (Dept. of Mathematics and Statistics, Univ. of Massachusetts, Amherst, MA 01002-4515)

A computer system is discussed that provides real-time accompaniment to a live musician playing a non-improvisatory piece of music. Particular attention is devoted to the ''listening'' process, in which the computer must follow the soloist's progress through the musical score by interpreting the sampled acoustic signal. The process is complicated by the significant variation and occasional errors from the live player during performance. A hidden Markov model is introduced providing a principled, trainable, and fast solution to the listening problem. The system is capable of assessing its own level of uncertainty about score position, as well as accommodating the sometimes strong signal component from the accompaniment instrument. A live demonstration will be provided.

11:00

2aMU6. Music scene description: Toward audio-based real-time music understanding. Masataka Goto (''Information and Human Activity,'' PRESTO, JST. / AIST (former ETL)., 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan, m.goto@aist.go.jp)

Music understanding is an important component of audio-based interactive music systems. A real-time *music scene description* system for the computational modeling of music understanding is proposed. This research is based on the assumption that a listener understands music without deriving musical scores or even fully segregating signals. In keeping with this assumption, our *music scene description* system produces intuitive descriptions of music, such as the beat structure and the melody and bass lines. Two real-time subsystems have been developed, a beat-tracking subsystem and a melody-and-bass detection subsystem, which can deal with real-world monaural audio signals sampled from popular-music CDs. The beat-tracking subsystem recognizes a hierarchical beat structure comprising the quarter-note, half-note, and measure levels by using three kinds of musical knowledge: of onset times, of chord changes, and of drum patterns. The melody-and-bass detection subsystem estimates the $F0$ (fundamental frequency) of melody and bass lines by using a predominant-$F0$ estimation method called *PreFEst*, which does not rely on the $F0$'s unreliable frequency component and obtains the most predominant $F0$ supported by harmonics within an intentionally limited frequency range. Several applications of music understanding are described, including a beat-driven, real-time computer graphics and lighting controller.

TUESDAY MORNING, 4 JUNE 2002 DUQUESNE ROOM, 9:00 TO 10:45 A.M.

Session 2aNS

Noise, Architectural Acoustics and Structural Acoustics and Vibration: Acoustics of Glazing and of Urban Spaces

Gregory C. Tocci, Chair
Cavanaugh Tocci Associates, 327F Boston Post Road, Sudbury, Massachusetts 01776

Chair's Introduction—9:00

Invited Papers

9:05

2aNS1. Analytic modeling of monolithic, laminated, and insulating glass sound transmission loss. Gregory Tocci (Cavanaugh Tocci Assoc., Inc., 327F Boston Post Rd., Sudbury, MA 01776)

The sound transmission loss of single lights of glass, and of insulating glass configurations using various combinations of monolithic and laminated glass, have been computed using Mathcad, and formulations of sound transmission loss as described by Beranek and Ver, and Au and Byrne. The importance of damping in these formulations, as seen by comparing estimated and measured TLs, is discussed. This discussion is extended to demonstrate the need for unifying methods for mounting glass and window samples in laboratory test openings, and to suggest ways to do so. [Work supported by Solutia, Inc.]

9:25

2aNS2. Sound isolation performance of interior acoustical sash. Gregory Tocci (Cavanaugh Tocci Assoc., Inc., 327F Boston Post Rd., Sudbury, MA 01776)

In existing, as well as new buildings, an interior light of glass mounted on the inside of a prime window is used to improve the sound transmission loss otherwise obtained by the prime window alone. Interior acoustical sash is most often 1/4 in. (6 mm) monolithic or laminated glass, and is typically spaced 3 in. to 6 in. from the glass of the prime window. This paper presents TL data measured at Riverbank Acoustical Laboratories by Solutia (formerly Monsanto) for lightweight prime windows of various types, with and without interior acoustical sash glazed with 1/4 in. laminated glass. The TL data are used to estimate the A-weighted insertion loss of interior acoustical sash when applied to prime windows glazed with lightweight glass for four transportation noise source types—highway traffic, aircraft, electric rail, and diesel rail. The analysis also has been extended to determine the insertion loss expressed as a change in OITC. The data also exhibit the reductions in insertion loss that can result from short-circuiting the interior acoustical sash with the prime window. [Work supported by Solutia, Inc.]

Contributed Papers

9:45

2aNS3. Practical problems of sound isolation through interior glazing systems. Carl Rosenberg and Dorie Najolia (Acentech, Inc., 33 Moulton St., Cambridge, MA 02138)

In a number of recent projects, designers have expanded the concepts of visual openness within a building to incorporate large expanses of glass. At the same time, building occupants expect the glass to achieve similar levels of acoustical isolation as stud constructed partitions. These situations include glass-enclosed conference rooms, clerestory lites between offices and corridors, curtain walls and strip windows, and vision panels that are adjacent to large atrium spaces. Variables impacting the acoustical performance include the size of the vision panels, the type of glass, the airspace between multiple layers, and the framing system that holds the glass. This paper presents a number of case studies and methodologies that address these issues.

10:00

2aNS4. The concept of privacy in the social practice of the living environment through sonic phenomena (case of the old city of Constantine. Algeria). Derbal Cobis Rabah and Zeghlache Hamza (BP.12.SMK. Constantine 25003. Algeria. cobis-derbal@iquebec.com)

The privacy of the different aspects composing the outside space and its social practice in the medinas and old islamic cities, is found to be often modulated by the quality of the physical criteria of the environment. Different activities in the medina have greatly modified the privacy surroundings. Privacy has been therefore, shaken up and greater permeability between inside and outside, particularly through sound phenomena, is one of the characteristics of such space. The exterior sound environment is fully perceived despite the supposed dispositions to preserve a certain degree of privacy. Underway studies on the old city of Constantine show already a great need to reconsider the concept of privacy in urban area throughout the sound environment. The noise acts, most of the time, not only by its disturbance, but by its information content as well, which is supposed to reduce and weaken significantly the concept of privacy in such old cities. More work is to be considered involving multidisciplinary approaches so that the notion of privacy can be worked out in reforming the discomfort of sonic environment into easy using soundscape with which privacy through social and historical mutations would fit in well.

10:15

2aNS5. Noise dimension in the design of urban spaces: Sonic instrumentation of public places. Derbal Cobis Rabah (BP. 12, SMK, Constantine 25003, Algeria, cobis-derbal@iquebec.com)

The sonic environment dimension is not often taken into account in the design of urban spaces. This is not the same for internal spaces inside which a certain control of the quality of the sonic conditions is henceforth possible. However in the case of urban spaces, one is often confronted with situations complex and difficult to handle. This always leads to an impredictible situation in terms of noise and urban sonic environment. Underway studies have shown the importance of the quality of the urban sonic environment in the daily use of public places. The latter, as elements that mark strongly the degree of dynamism of the urban structure of our cities, are often subjected to a very complex situation of noise pollution. Introducing a source of noise that might have an effect of recomposing positively the sonic environment of the public place was the main approach of the study. It can be used as well to restore the sound in its utilitarian and practical function and to articulate the perception of the spatial and temporal aspects of the place. It is a sort of a sonic instrumentation of urban spaces that might recompose or recharacterize urban and architectural environments.

10:30

2aNS6. A multidisciplinary approach of the problem of noise nuisance in urban environment. Derbal Cobis Rabah and Zeghlache Hamza (BP. 12, SMK, Constantine 25003, Algeria, cobis-derbal@iquebec.com)

More often the problem of noise and sonic pollution, particularly in urban sites, is studied by different disciplines such as physics, the acoustics, the psychoacoustics, the medicine and others. It is independently of each other that these sciences are often approaching this subject. Some studies are carried out in laboratories taking noise as samples cut off their realistic context. Urban noise is studied as well by making an abstraction of the different contextual parameters by idealizing a rather complex sonic environment. The noise, according to this present approach, is suposed to react with surounding space, and it takes the form and the quality of the place by defining and requalifying it. It is found that the contextual aspects such as social, cultural or even symbolic dimensions modulate the listening conditions and the perception quality of the noise and even the living and the daily practice of the urban space. The multiparameter dimension study of the noise in an urban context is necessary to better work out the problem and to try to come up with some practical and efficient solutions. The little amount of studies based on such multidisciplinary approach, confort well our effort to go ahead with this methodological approach.

Session 2aPA

Physical Acoustics and Underwater Acoustics: Recent Developments in Wave Propagation in Random and Complex Media I

D. Keith Wilson, Chair
AMSRL CI EE, U.S. Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland 20783-1197

Chair's Introduction—8:25

Invited Papers

8:30

2aPA1. Quasiwavelet models of sound scattering by atmospheric turbulence. George H. Goedecke (Dept. of Phys., New Mexico State Univ., Las Cruces, NM 88003, ggoedeck@nmsu.edu), Vladimir E. Ostashev (New Mexico State Univ., Las Cruces, NM 88003 and NOAA/Environ. Tech. Lab., Boulder, CO), D. Keith Wilson, and Harry J. Auvermann (U.S. Army Res. Lab., Adelphi, MD)

Quasiwavelet (QW) representations of turbulence are composed of self-similar, localized, eddylike structures. The QW functions are not true wavelets, in that they do not form a mathematically complete basis or have zero mean. Nevertheless, they appear to be very useful for applications involving scattering and propagation of sound waves. In this paper, the QW formulation of Goedecke and Auvermann [J. Acoust. Soc. Am. **102**, 759–771 (1997)] is outlined. The QW expressions for the spatial spectra and the corresponding sound scattering cross sections due to the velocity and temperature fluctuations of isotropic homogeneous turbulence are discussed. The spectra for different eddy structures are always similar to the von Karman spectra, and agree with the Kolmogorov spectra in the inertial range. Equations that yield the QW eddy functions in terms of the spectra are derived, and a QW function is found that yields the von Karman velocity spectrum exactly. Some results are presented from a numerical calculation of coherent scattering and temporal spectral broadening due to advecting turbulence modeled by QW eddies flowing with a wind. Future applications to modeling scattering by anisotropic and/or inhomogeneous turbulence are discussed. [Work supported by the ARO under Contract No. DAAD19-01-1-0640 (administered by W. Bach).]

8:50

2aPA2. Some validity issues in the theory and modeling of WPRM. Terry Ewart and Frank Henyey (Univ. of Washington, M.S. 355640, Seattle, WA 98105)

In many publications in the recent to older litererature on the theory and modeling of wave propagation in random media, WPRM, in ocean and atmospheric environments there are serious validity issues that arise. The issues discussed will include the following. (1) The validity of using the Markov approximation in theoretical treatments must be established for specific environments before theories based on that approximation can be used. Parameters that are required to be small must be shown to be so before the applicabilty to specific environments can be assured. (2) Most publications that purport to model the pdf's of intensity do not include the standard ''goodness of fit'' parameters. The Kolmogorov–Smirnov GoF test emphasizes the center of the distribution and often the pdf's of WPRM are characterized by very high tailed distributions where the fitting of that component can be important to understanding the scattering physics. GoF tests that include the full range of received intensity are needed. (3) Signal processors invariably require that the quadrature components of their receptions have Gaussian pdf's. It has been shown by simulations of WPRM that this is often not even close to the case. What significance this effect has on signal processing algorithms is a difficult and unresolved issue.

9:10

2aPA3. Recent results from basin-scale ocean acoustic wave propagation: Implications for the general theory of the approach to saturation. John A. Colosi (Woods Hole Oceanogr. Inst., M.S. 11, Woods Hole, MA 02543)

Long-range ocean acoustics as a problem in wave propagation through random media is unique in several ways. First the propagation takes place in the ocean sound channel (a waveguide), and second, a primary source of sound speed fluctuations are the anisotropic and inhomogeneous ocean internal waves (not turbulence). Recent observations of 75 Hz, broadband transmissions over several thousand kilometers have revealed fascinating results on the approach to saturation. Two distinct propagation regimes are observed in the same pulse: a fully saturated regime characterized by Rayleigh statistics, and an unsaturated regime in which the probability density function for intensity is nearly log-normal. The transition between these two regimes is extremely rapid. The saturated region of the pulse is composed of acoustic energy traveling near the waveguide axis, while the unsaturated region is composed of acoustic energy strongly refracted by the waveguide. This fact suggests a fundamental stabilizing influence by the waveguide. Application of path-integral theory, which was successful in describing shorter-range ocean experiments, cannot describe these two regimes. Recent efforts using semiclassical and ray chaos theory have led to new insights into the observational results and to the failure of the path-integral theory, yet a complete theoretical description remains elusive.

9:30–9:45 Break

9:45

2aPA4. An eigenfunction expansion for low-frequency acoustic propagation in a downward-refracting stratified medium over a complex impedance plane. Roger Waxler (NCPA, Coliseum Dr., University, MS 38677-1848, rwax@olemiss.edu)

The well-known eigenfunction expansions associated with self-adjoint operators cannot be used to describe sound propagation over a complex impedance plane since, if the impedance has a real part, the resulting impedance boundary condition is not self-adjoint. Instead, a less widely known eigenfunction expansion associated with non-self-adjoint operators can be used. This eigenfunction expansion is applied to a vertically stratified model for downward refraction in the nighttime boundary layer. As in the self-adjoint case the propagation separates into a ducted part, expressed as a sum of modes which decay exponentially with height, and an upwardly propagating part, expressed as an integral over modes which are asymptotically (with height) plane waves. The eigenvalues associated with the ducted modes are complex, the imaginary parts being related to the acoustic attenuation. An efficient method for finding the complex eigenvalues is introduced from which a physically intuitive form of the attenuation coefficients is obtained. For low-frequency propagation (100 Hz or less) the number of modes is small, making this a simple way to model the ducted part of the propagation.

10:00

2aPA5. Including attenuation and dispersion in time domain modeling of broadband sound propagation in dispersive oceanic media. Guy V. Norton (Naval Res. Lab., Stennis Space Center, MS 39529-5004, norton@nrlssc.navy.mil) and Jorge C. Novarini (Planning Systems, Inc., Long Beach, MS 39560)

Shallow water sound propagation almost invariably involves paths through regions occupied by assemblages of microbubbles near the air/water interface and through the underlying sediment. Such an environment is intrinsically dispersive. Any reliable numerical model of acoustic propagation in this type of environment requires the inclusion of attenuation and its causal companion, dispersion. The time domain is the natural domain for numerical modeling of broadband pulse propagation. The inclusion of causal attenuation and dispersion within a numerical time domain model has in the past presented a challenge to modelers. For the case of propagation in a linear medium, Szabo [T. L. Szabo, J. Acoust. Soc. Am. **96**, 491–500 (1994)], making use of generalized functions, introduced the concept of a convolutional propagation operator that plays the role of a causal propagation factor in the time domain. Utilizing a Finite Difference Time Domain (FDTD) model it is shown that the inclusion of the convolutional propagation operator correctly carries the information of attenuation and dispersion into the time domain. The causal FDTD model was exercised on environments having functional forms for the attenuation of the type encountered at sea, both in typical bubble clouds as well as the underlying sediment. [Work supported by ONR/NRL.]

10:15

2aPA6. Ultrasonic b-scan imaging with adaptive beamformation using aberration correction. Wayne C. Pilkington (Dept. of Elec. and Computer Eng., Univ. of Rochester, Rochester, NY 14627, pilkingt@ece.rochester.edu), James C. Lacefield (Depts. of Elec. and Computer Eng. and Medical Biophys., Univ. of Western Ontario, London, ON N6A 5B9, Canada), and Robert C. Waag (Depts. of Elec. and Computer Eng. and Radiol., Univ. of Rochester, Rochester, NY 14627)

The effectiveness of adaptive beam formation using aberration correction has been demonstrated in ultrasonic b-scans of liver-mimicking scattering phantoms imaged with and without an intervening aberrator that produced wavefront distortion comparable to that of abdominal wall. Images of 4 mm diam spherical features (either positive or negative contrast lesions or scatterer-free cysts) in the uniform scattering background of the phantoms were produced at 3.0 MHz with a two-dimensional (80×80-element) array transducer system. Time-shift aberration was estimated from the scattering data and used to compensate both transmit and receive waveforms. Image improvements were assessed by comparison of feature contrast with and without aberration correction in individual images and by comparison of intensities in averages of independent, statistically identical images. Feature contrasts and borders were visibly and measurably improved, sometimes to near the water path results, using aberration correction, particularly when both transmit and receive corrections were applied. An efficient implementation of aberration correction was achieved by correction of multiple image scan lines with a single aberration estimate. Aberration correction using estimates from one-seventh the number of scan lines in 8 mm wide images produced improvements comparable to those achieved by individually estimating and correcting aberration in every scan line.

10:30

2aPA7. Statistical estimation of propagation path parameters for ultrasonic aberration correction. Robert C. Waag (Depts. of Elec. and Computer Eng. and Radiol., Univ. of Rochester, Rochester, NY 14627) and Jeffrey P. Astheimer (66 Sibley Rd., Honeoye Falls, NY 14472)

Coefficients of a finite impulse response linear filter model for ultrasonic propagation are found using a statistical estimation. The model employs a Green's function to describe scattering received by a two-dimensional array from a random-medium volume illuminated by a transmit beam. The frequency response of this Green's function is factored into a homogeneous-transmission term and a path-dependent aberration term. Relative amplitude and phase of the aberration response are estimated over the frequency band of the transmit–receive system by using scattering from closely situated volumes and assuming that the aberration is the same for each volume. The amplitude and phase are obtained from the power and cross-power spectra of signals at positions throughout the receive aperture. The homogeneous response is estimated by averaging and is removed to isolate the aberration response. Propagation path aberration parameters calculated from pulse-echo measurements of random scattering through a tissue-mimicking aberration phantom are similar to corresponding parameters calculated for the same aberrator and array position by using echoes from a point-like reflector. The results indicate the approach is capable of describing, in addition to time-shifts, changes in waveform amplitude and shape produced by propagation through a distributed aberration.

10:45

2aPA8. Simulation of ultrasonic focus aberration and correction through human tissue. Makoto Tabei (Dept. of Elec. and Computer Eng., Univ. of Rochester, Rochester, NY 14627), T. Douglas Mast (Ethicon Endo-Surgery, 4545 Creek Rd., ML 40, Cincinnati, OH 45242), and Robert C. Waag (Univ. of Rochester, Rochester, NY 14627)

Ultrasonic focusing in two dimensions has been investigated by calculating the propagation of ultrasonic pulses through cross-sectional models of the human abdominal wall and breast. Propagation calculations used a full-wave k-space method that accounts for spatial variations in density, sound speed, and frequency-dependent absorption and includes perfectly matched layer absorbing boundary conditions. To obtain a distorted receive wavefront, propagation from a point source through the tissue path was computed. Receive focusing used an angular spectrum method. Transmit focusing was accomplished by propagating a pressure wavefront from a virtual array through the tissue path. As well as uncompensated focusing, focusing that employed time-shift compensation and time-shift compensation after backpropagation was investigated in both transmit and receive and time reversal was investigated for transmit focusing in addition. The results indicate, consistent with measurements, that the breast causes greater focus degradation than the abdominal wall. The investigated compensation methods corrected the receive focus better than the transmit focus. Time-shift compensation after backpropagation improved the focus from that obtained using time-shift compensation alone but the improvement was less in transmit focusing than in receive focusing. Transmit focusing by time reversal resulted in lower sidelobes but larger mainlobes than the other investigated transmit focus compensation methods.

Session 2aPPa

Psychological and Physiological Acoustics: Hearing Aid Design: Psychophysics and Signal Processing I

Brent W. Edwards, Cochair
Sound ID, 3430 West Bayshore Road, Palo Alto, California 94303

Dianne J. Van Tasell, Cochair
Starkey Laboratories, 6600 Washington Avenue, South, Eden Prairie, Minnesota 55344

Invited Papers

8:30

2aPPa1. State of the art in perceptual design of hearing aids. Brent W. Edwards (Sound ID, 3430 W. Bayshore Rd., Palo Alto, CA 94303, brent@edwards.net) and Dianne J. Van Tasell (Starkey Labs., Eden Prairie, MN 55344)

Hearing aid capabilities have increased dramatically over the past six years, in large part due to the development of small, low-power digital signal processing chips suitable for hearing aid applications. As hearing aid signal processing capabilities increase, there will be new opportunities to apply perceptually based knowledge to technological development. Most hearing loss compensation techniques in today's hearing aids are based on simple estimates of audibility and loudness. As our understanding of the psychoacoustical and physiological characteristics of sensorineural hearing loss improves, the result should be improved design of hearing aids and fitting methods. The state of the art in hearing aids will be reviewed, including form factors, user requirements, and technology that improves speech intelligibility, sound quality, and functionality. General areas of auditory perception that remain unaddressed by current hearing aid technology will be discussed.

8:45

2aPPa2. It's not your father's recruitment: A new view of loudness growth in cochlear hearing loss. Mary Florentine (Dept. of Speech-Lang. Path. & Audiol. (133 FR), Northeastern Univ., 360 Huntington Ave., Boston MA 02115. florentin@neu.edu) and Søren Buus (Comm. & Dig. Sig. Proc. Ctr., Dept. of Elec. & Computer Eng. (440 DA), Northeastern Univ., 360 Huntington Ave., Boston, MA 02115)

For over 60 years, one of the most commonly held beliefs in audiology has been that loudness grows more rapidly than normal near the elevated thresholds of listeners with cochlear hearing losses. This belief was based on the untested assumption that loudness at threshold is the same—often incorrectly assumed to be zero—in normal listeners and listeners with cochlear hearing losses. Recent tests of this assumption indicate that loudness at threshold is not zero in either group of listeners. These data also indicate that loudness near threshold grows at a normal rate in listeners with cochlear hearing losses. In other words, recruitment in the sense of a rapid growth of loudness near an elevated threshold does not occur. How, then, can intense sounds have approximately normal loudness in listeners with cochlear hearing losses? The explanation is that loudness at threshold is greater when the threshold is elevated by a cochlear loss than when it is normal. We call this *softness imperception*. Therefore, listeners with cochlear hearing losses not only have reduced dynamic range of audibility; they also have reduced dynamic range of loudness. Implications for the design of hearing aids will be discussed. [Work supported by NIH/NIDCD grant R01DC02241.]

9:15

2aPPa3. Behavioral estimates of cochlear nonlinearity and its effects on normal and impaired hearing. Andrew J. Oxenham (Res. Lab of Electron., MIT, Cambridge, MA 02139) and Christopher J. Plack (Univ. of Essex, Wivenhoe, Colchester CO4 3SQ, UK)

Recent physiological studies of basilar-membrane motion have clarified many aspects of normal and pathological cochlear processing. The compressive input–output function in response to tones around the characteristic frequency and the sharp tuning at low levels are examples of basilar-membrane properties that are thought to be important for hearing, but that are also highly vulnerable to cochlear damage. Aspects of auditory perception influenced by cochlear nonlinearity include loudness and dynamic range, temporal processing, and frequency selectivity. Functional models have assisted us in understanding the perceptual consequences of peripheral nonlinearity in a wide variety of psychoacoustic tasks. In addition, many effects of cochlear hearing loss can be simulated within such a model simply by reducing or eliminating the nonlinearity within the model. A number of behavioral measures of cochlear nonlinearity in humans have been developed in recent years. These techniques offer insights into human cochlear

processing in general, and may also provide diagnostic information about cochlear function on an individual basis. By gaining a better understanding of the changes in cochlear processing associated with hearing loss, it may be possible to design sound-processing algorithms for hearing aids that better compensate for the effects of cochlear damage. [Work supported by NIH Grant R01DC03909.]

9:45–10:00 Break

10:00

2aPPa4. Should visual speech cues (speechreading) be considered when fitting hearing aids? Ken Grant (Walter Reed Army Medical Ctr., Army Audiol. and Speech Ctr., Washington, DC 20307-5001)

When talker and listener are face-to-face, visual speech cues become an important part of the communication environment, and yet, these cues are seldom considered when designing hearing aids. Models of auditory–visual speech recognition highlight the importance of complementary versus redundant speech information for predicting auditory–visual recognition performance. Thus, for hearing aids to work optimally when visual speech cues are present, it is important to know whether the cues provided by amplification and the cues provided by speechreading complement each other. In this talk, data will be reviewed that show nonmonotonicity between auditory-alone speech recognition and auditory–visual speech recognition, suggesting that efforts designed solely to improve auditory-alone recognition may not always result in improved auditory–visual recognition. Data will also be presented showing that one of the most important speech cues for enhancing auditory–visual speech recognition performance, voicing, is often the cue that benefits least from amplification.

10:30

2aPPa5. Physiological modeling for hearing aid design. Ian C. Bruce,[a] Eric D. Young, and Murray B. Sachs (Ctr. for Hearing Sci. and Dept. of Biomed. Eng., Johns Hopkins Univ., 720 Rutland Ave., Baltimore, MD 21205, ibruce@ieee.org)

Physiological data from hearing-impaired cats suggest that conventional hearing aid signal-processing schemes do not restore normal auditory-nerve responses to a vowel [Miller *et al.*, J. Acoust. Soc. Am. **101**, 3602 (1997)] and can even produce anomalous and potentially confounding patterns of activity [Schilling *et al.*, Hear. Res. **117**, 57 (1998)]. These deficits in the neural representation may account at least partially for poor speech perception in some hearing aid users. An amplification scheme has been developed that produces neural responses to a vowel more like those seen in normal cats and that reduces confounding responses [Miller *et al.*, J. Acoust. Soc. Am. **106**, 2693 (1999)]. A physiologically accurate model of the normal and impaired auditory periphery would provide simpler and quicker testing of such potential hearing aid designs. Details of such a model, based on that of Zhang *et al.* [J. Acoust. Soc. Am. **109**, 648 (2001)], will be presented. Model predictions suggest that impairment of both outer- and inner-hair cells contribute to the degraded representation of vowels in hearing-impaired cats. The model is currently being used to develop and test a generalization of the Miller *et al.* speech-processing algorithm described above to running speech. [Work supported by NIDCD Grants DC00109 and DC00023.] [a] Now with the Dept. of Electrical and Computer Engineering, McMaster Univ., 1280 Main St. W., Hamilton, ON L8S 4K1, Canada.

11:00

2aPPa6. Biologically inspired binaural hearing aid algorithms: Design principles and effectiveness. Albert Feng (Beckman Inst., Univ. of Illinois, Urbana, IL 61801)

Despite rapid advances in the sophistication of hearing aid technology and microelectronics, listening in noise remains problematic for people with hearing impairment. To solve this problem two algorithms were designed for use in binaural hearing aid systems. The signal processing strategies are based on principles in auditory physiology and psychophysics: (a) the location/extraction (L/E) binaural computational scheme determines the directions of source locations and cancels noise by applying a simple subtraction method over every frequency band; and (b) the frequency-domain minimum-variance (FMV) scheme extracts a target sound from a known direction amidst multiple interfering sound sources. Both algorithms were evaluated using standard metrics such as signal-to-noise-ratio gain and articulation index. Results were compared with those from conventional adaptive beam-forming algorithms. In free-field tests with multiple interfering sound sources our algorithms performed better than conventional algorithms. Preliminary intelligibility and speech reception results in multitalker environments showed gains for every listener with normal or impaired hearing when the signals were processed in real time with the FMV binaural hearing aid algorithm. [Work supported by NIH-NIDCD Grant No. R21DC04840 and the Beckman Institute.]

Session 2aPPb

Psychological and Physiological Acoustics: Potpourri (Poster Session)

John G. Neuhoff, Chair
Psychology Department, College of Wooster, Wooster, Ohio 44691

Contributed Papers

All posters will be on display from 8:30 a.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 12 noon. To allow for extended viewing time, posters will remain on display until 5:00 p.m.

2aPPb1. A cortical network underpinning the perceptual priority for rising intensity and auditory "looming." John G. Neuhoff (Dept. of Psych., The College of Wooster, Wooster, OH 44691, jneuhoff@wooster.edu), Deniz Bilecen, Henrietta Mustovic, Hartmut Schachinger, Erich Seifritz (Univ. of Basel, Basel, Switzerland), Klaus Scheffler (Univ. of Freiburg, Freiburg, Germany), and Francesco Di Salle (Univ. of Naples Federico II, Naples, Italy)

Relative motion between a sound source and a listener creates a change in acoustic intensity that can be used to anticipate the source's approach. Humans have been shown to overestimate the intensity change of rising compared to falling intensity sounds and underestimate the time-to-contact of approaching sound sources. From an evolutionary perspective, this perceptual priority for looming sounds may represent an adaptive advantage that provides an increased margin of safety for responding to approaching auditory objects. Here, using functional magnetic resonance imaging, we show that the prioritization of rising contrasted with falling intensity sine-tones is grounded in a specific neural network. This network is predominantly composed of the superior temporal sulci, the middle temporal gyri, the right temporo-parietal junction, the motor and premotor cortices mainly on the right hemisphere, the left frontal operculum, and the left superior posterior cerebellar cortex. These regions are critical for the allocation of attention, the analysis of space, object recognition, and neurobehavioral preparation for action. Our results identify a widespread neural network underpinning the perceptual priority for looming sounds that can be used in translating sensory information into preparedness for adverse events and appropriate action. [Work supported by the Swiss and the American NSFs.]

2aPPb2. Effects of spatially correlated acoustic–tactile information on judgments of auditory circular direction. Annabel J. Cohen, M. J. Reina Lamothe, Ian D. Toms, and Richard A. G. Fleming (Dept. of Psych., Univ. of PE, Charlottetown, PE C1A 4P3, Canada)

Cohen, Lamothe, Fleming, MacIsaac, and Lamoureux [J. Acoust. Soc. Am. **109**, 2460 (2001)] reported that proximity governed circular direction judgments (clockwise/counterclockwise) of two successive tones emanating from all pairs of 12 speakers located at 30-degree intervals around a listeners' head (cranium). Many listeners appeared to experience systematic front–back confusion. Diametrically opposed locations (180-degrees—theoretically ambiguous direction) produced a direction bias pattern resembling Deutsch's tritone paradox [Deutsch, Kuyper, and Fisher, Music Percept. **5**, 7992 (1987)]. In Experiment 1 of the present study, the circular direction task was conducted in the tactile domain using 12 circumcranial points of vibration. For all 5 participants, proximity governed direction (without front–back confusion) and a simple clockwise bias was shown for 180-degree pairs. Experiment 2 tested 9 new participants in one unimodal auditory condition and two bimodal auditory–tactile conditions (spatially-correlated/spatially-uncorrelated). Correlated auditory–tactile information eliminated front–back confusion for 8 participants and replaced the ''paradoxical'' bias for 180-degree pairs with the clockwise bias. Thus, spatially correlated audio–tactile location information improves the veridical representation of 360-degree acoustic space, and modality-specific principles are implicated by the unique circular direction bias patterns for 180-degree pairs in the separate auditory and tactile modalities. [Work supported by NSERC.]

2aPPb3. Localization judgments are sensitive to late-arriving sound. G. Christopher Stecker (Kresge Hearing Res. Inst., Univ. of Michigan, 1301 E. Ann St., Ann Arbor, MI 48109-0519, cstecker@umich.edu) and Ervin R. Hafter (Univ. of California at Berkeley, Berkeley, CA 94720-1650)

In a study of temporal aspects of sound localization, free-field listeners localized trains of 2–32 spatially distributed clicks (4-kHz narrow-band impulses) in a pointing task. Multiple regression coefficients estimated the perceptual ''weight'' listeners applied to each click in a train when making location judgments. Temporal weighting functions (TWFs) obtained in this way exhibited two dominant features: First, at high stimulus rates (>80 Hz), weight for the first click in a train was significantly larger than that for later clicks, as observed previously in studies measuring TWFs for ITD-based lateralization [K. Saberi, Percept. Psychophys. **58**, 1037–1046 (1998); R. Dizon *et al.*, Assoc. Res. Otolaryngol. Abs. **21**, 42 (1998)]. Second, in several conditions, clicks near the end of a train received larger weights than earlier clicks. This presentation focuses on the latter finding, which affected the last few clicks in a train despite manipulations of train length (overall duration) and predictability, and which appeared independently of increased weight on the first click. The effect cannot be explained as a release from temporary suppression, but may relate to effects of temporal integration, apparent motion, or the influence of non-ITD spatial cues. [Work supported by NIDCD 00087.]

2aPPb4. Simple neuron models of ITD sensitive neurons. Vasant Dasika (Hearing Res. Ctr. and Dept. of Biomed. Eng., Boston Univ., Boston, MA 02215), John A. White, and H. Steven Colburn (Boston Univ., Boston, MA)

Neurons which show sensitivity to interaural time delay (ITD) exist in both mammalian medial superior olive (MSO), and bird nucleus laminaris (NL). In this study, we examine simple mathematical models of single MSO and NL cells which respond probabilistically to a pair of isolated inputs with a response probability that depends on the input interpulse interval. Inputs are either isolated pulse pairs or pairs of periodic trains, with or without random jitter added to their event times. Refractoriness is incorporated in the input description and/or in the cell model in specified simulations. We find that periodic rate-ITD shapes are shaped by three

interacting factors: the cell's temporal response (described by the paired-pulse response), input frequency, and the degree of input synchrony. Paired-pulse responses are able to predict the widths of rate-ITD curves obtained from deterministic periodic input simulations. Reduced input synchrony predictably smears rate-ITD curves. Larger numbers of weaker inputs yield stronger rate-ITD modulation than a few strong inputs. Model response is compared with *in vivo* and *in vitro* MSO and NL physiological data. Comparisons with published analytical models as well as more complex and realistic physiological cell models are examined.

2aPPb5. JNDS of interaural time delay (ITD) of selected frequency bands in speech and music signals. Avner Aliphas, H. Steven Colburn (Hearing Res. Ctr. and Dept. of Biomed. Eng., Boston Univ., Boston, MA 02215), and Oded Ghitza (Agere Systems)

JNDS of interaural time delay (ITD) of selected frequency bands in the presence of other frequency bands have been reported for noiseband stimuli [Zurek (1985); Trahiotis and Bernstein (1990)]. Similar measurements will be reported for speech and music signals. When stimuli are synthesized with bandpass/band-stop operations, performance with complex stimuli are similar to noisebands (JNDS in tens or hundreds of microseconds); however, the resulting waveforms, when viewed through a model of the auditory periphery, show distortions (irregularities in phase and level) at the boundaries of the target band of frequencies. An alternate synthesis method based upon group-delay filtering operations does not show these distortions and is being used for the current measurements. Preliminary measurements indicate that when music stimuli are created using the new techniques, JNDS of ITDs are increased significantly compared to previous studies, with values on the order of milliseconds.

2aPPb6. Trajectory perception in the free field. Joshua A. Miele and Ervin R. Hafter (Dept. of Psych., Univ. of California, Berkeley, CA 94720, jam@socrates.berkeley.edu)

Using virtual sound sources in an echo-attenuated chamber, the sensitivity of subjects to the direction of linear auditory motion was evaluated. Four sighted and four blind subjects participated. Subjects reported the perceived trajectory by orienting a freely rotating pointer. Subjects fell into two distinct categories: those who performed well at the task, and those who did not, with blind subjects tending to be in the first category and sighted subjects the second. Those subjects performing well at the task also showed improved performance with increasing linear velocity. Because the only stimulus parameter associated with increased linear velocity was frequency envelope, it can be concluded that frequency changes resulting from the Doppler effect can improve the accuracy of auditory trajectory perception. The same subjects performed discrimination tasks measuring sensitivity to changes in azimuthal acceleration, as well as to changes in amplitude and frequency envelopes. Two models of trajectory perception are considered: A Psychoacoustical Parameters Model uses a weighted combination of specific perceptual abilities such as sensitivity to azimuthal acceleration, and to changes in amplitude and frequency envelopes. An Informational Masking Model is considered as a possible explanation for those subjects demonstrating poor performance in the trajectory perception task.

2aPPb7. Localization of noise bursts in the median plane: Effects of duration and level. Gongqiang Yu, Ruth Y. Litovsky, and Frederic L. Wightman (Waisman Ctr., Univ. of Wisconsin-Madison, Madison, WI 53705, yu@waisman.wisc.edu)

Previous research has revealed that human sound localization in the median sagittal and in the front/back dimension may depend on the listener's ability to reliably extract information regarding characteristic peaks and notches of the source spectrum upon reaching the ears. While long-duration sounds are well localized using spectral cues, the processing of spectral cues for sounds of short duration appears to be degraded. The mechanisms involved in ''image formation'' in the localization process, and the relative importance of sound duration and level are not well understood. In the present study, the effect of signal duration and level on localization performance, with an emphasis on the median-sagittal plane were systematically explored. Three types of stimuli, clicks, noises, and trains of frozen and random noise bursts encompassing the same time period as the noises, with different duration and at various levels, were randomly presented to the loudspeakers. Subjects were blindfolded, but were instructed to move their heads as necessary, to maximize their confidence regarding source position judgment. Results will be discussed in relation to possible mechanisms employed by the auditory system to utilize spectrally shaped cues in the localization process. [Work supported by NIDCD Grant Nos. R29 DC03083 and P01 DC 00116.]

2aPPb8. Spectral–temporal analysis of OAEs in normal-hearing humans: Latencies, fine structure, and multiple internal reflections. Dawn Konrad-Martin (Dept. of Commun. Disord. and Sci., Rush Univ., 1653 W. Congress Pkwy., Rm. 1015 AAC, Chicago, IL 60612), Douglas H. Keefe, and Jeffrey L. Simmons (Boys Town Natl. Res. Hospital, Omaha, NE 68131)

Otoacoustic emissions (OAEs) display complexity in both time and frequency, which may be obscured in traditional response analyses. Time-frequency representations (TFR) of the eliciting stimuli and the recorded OAE provide a tool to understand this complexity. TFRs were applied to stimulus-frequency (SF) and distortion-product (DP) OAEs obtained in 25 normal-hearing adults, using three classes of stimuli: continuous tones, gated tones (tonal stimuli with well-defined onset, steady state and decay), and tone pips (band-limited impulses). Synchronous spontaneous OAEs were measured using clicks to assess their contributions to TFRs of SFOAE and DPOAEs. A common form of TFR generated from tone-pips is a collection of frequency-specific components, with each component characterized by one or more brief segments or a single long-duration segment. Frequency-pulling effects between components were observed, however, little evidence was found for interactions within transient SFOAEs in the form of intermodulation distortion. TFRs show evidence for strong emission regions or cochlear ''hot spots'' within pip responses, and multiple internal reflections within gated and pip responses. The resulting measures of OAE latencies are compared with model predictions. Thus, the TFR analysis provides a quantitative tool for measuring the spectral-temporal organization of multiple-component OAEs. [Work supported by NIH, DC003784.]

2aPPb9. SFOAE input/output functions elicited by slightly off-frequency suppressors. Kim S. Schairer and Douglas H. Keefe (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131, schairerk@boystown.org)

Basilar membrane (BM) response growth is thought to be related to stimulus frequency otoacoustic emission (SFOAE) input/output (I/O) functions. SFOAEs comprise the response to sinusoidal excitation and provide an estimate of nonlinearity at a tonotopic place on the BM. SFOAE I/O functions at frequency f_0 (across levels L_0) are measured in the presence of suppressor f_s (at level L_s) in subjects with normal hearing. f_s is varied below, equal to, and above f_0 in the range $0.93 < f_s/f_0 < 1.07$. For fixed suppressor level ($L_s = 80$ dB SPL) and varying $L_0 < L_s$, the SFOAE I/O function is predicted to be closely related to the BM I/O function. Measured I/O functions are approximately invariant as f_s is varied. Functions saturate at increasing levels with a decrease in some ears at the highest levels. An alternative paradigm in which $L_s = L_0$ are covaried yields SFOAE I/O functions with local minima, but which may have a less direct connection to BM response growth. Based on primaries at f_s and f_0, SFOAE and DPOAE I/O functions differ. One implication is that SFOAE I/O functions may be used to relate BM nonlinear response growth to perceptual processes, including loudness growth and forward masking. [Work supported by NIH (R01 DC003784, T32 DC00013).]

2aPPb10. The effect of ear canal pressure on pure-tone thresholds and middle-ear reflectance and conductance. M. Patrick Feeney, Julie Hazelbaker, J. Kip Kelly, and Amanda Graves (Speech and Hearing Sci., Ohio State Univ., Columbus, OH 43210)

The effect of ear canal pressure on pure-tone thresholds and wideband reflectance and conductance was examined in ten human participants. Pure-tone thresholds at 0.5, 1, 2, and 4 kHz were measured in random order using a 2AFC procedure at ambient pressure, and at ear canal pressures of positive and negative 200 daPa. These results were compared to wideband changes in reflectance and conductance using the same probe. Reflectance changes with pressure were in good agreement with recent data on reflectance tympanometry [R. H. Margolis, G. L. Saly, and D. H. Keefe, J. Acoust. Soc. Am. **106**, 265–280 (1999)]. Both one-third-octave conductance measures and pure-tone thresholds were found to decrease by around 10 dB at 500 Hz and 1000 Hz compared to a 2- to 4-dB decrease at 2000 and 4000 Hz. These results are consistent with the change in middle-ear sound transmission with ear-canal pressure changes reported in a recent temporal-bone study in Norwegian cattle [M. Kringlebotn, J. Acoust. Soc. Am. **107**, 1442–1450 (2000)]. The results of this study suggest that the change in middle-ear conductance may provide an objective measure of conductive hearing loss for the conditions in this experiment. [Work supported by NIH, RO3 DC04129.]

2aPPb11. Vibrotactile suppression of tinnitus. Martin L. Lenhardt (Prog. in Biomed. Eng. and Dept. of Otolaryngol., Virginia Commonwealth Univ., Richmond, VA 23298-0168)

At the Society's 142nd meeting, the efficacy of high frequency bone conducted stimulation in suppressing tinnitus was presented. The hypothesized mechanism was the reprogramming of frequency tuning of auditory neurons in the central nervous system, secondarily to peripheral hearing loss. This mechanism is unlikely in cases of tinnitus in the presence of normal audiometric sensitivity. There is the possibility that hearing loss above 10 kHz can play a role in tinnitus, an association not thoroughly explored. Somatomotor stimulation influencing the quality of tinnitus has been reported, as have interconnections of the auditory and somatosensory systems. There would appear to be an evolutionary advantage of linking the sensorimotor organization of the external ear and the auditory function of the brainstem in sound localization. Thus, stimulation of the pinna and post auricular area may be a means of suppressing tinnitus. To that end a thin aluminum ceramic bimorph was constructed to fit on the inner surface of the pinna. When driven by low ($<$100 Hz) and high ($>$10 kHz) frequencies multiplied by MHz carriers, demodulation in the skin resulted in vibrotactile stimulation. Tactile stimulation was an adjunct to the high frequencies resulting in a multimodal suppressive effect in a small pilot study.

2aPPb12. On the perceptual and behavioral significance of saccular acoustic sensitivity. Neil P. M. Todd (Dept. of Psych., Univ. of Manchester, Manchester M13 9PL, UK)

Evidence accumulated over the last few decades supports the case that the sacculus has conserved an acoustic sensitivity throughout vertebrate phylogeny [N. P. Todd, J. Acoust. Soc. Am. **110**, 380–390 (2001)]. In humans saccular acoustic sensitivity may be demonstrated by means of myogenic vestibular evoked potentials (MVEP) but the question remains whether saccular acoustic sensitivity in humans has any perceptual or behavioral significance. It has been shown that MVEP have the following properties: (1) MVEP can be obtained to natural acoustic stimuli above about 90 dB SPL; (2) MVEP have a frequency tuning property with a best frequency between 300 Hz–350 Hz; (3) MVEP exhibit adaptation, characteristic or normal sensory processing; and (4) there is a change of the quality of sensation above the MVEP threshold, indicative that saccular acoustic sensitivity may play a role in the perception of loud sound. In this paper new evidence is discussed which indicates that both myogenic and neurogenic vestibular evoked potentials (NVEP), particularly those obtainable from prefrontal cortex, may be linked to subjective and behavioral responses to loud sound. These responses may in turn be related to individual differences in the functioning of the mesolimbic dopamine system, including prefrontal cortex.

2aPPb13. A chinchilla nonlinear cochlear filterbank. Alberto Lopez-Najera, Enrique A. Lopez-Poveda (Centro Regional de Investigación Biomédica, Facultad de Medicina, Universidad de Castilla, La Mancha, 02071 Albacete, Spain), and Ray Meddis (Univ. of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK)

A dual-resonance nonlinear (DRNL) filter [Meddis *et al.,* J. Acoust. Soc. Am. **106**, 2852–2861 (2001)] was fitted to model chinchilla cochlear responses to tonal stimuli at individual sites along the basilar membrane (BM) with best frequencies (BF) of 0.8, 5.5, 7.25, 9.75, 10.0, 12.0, and 14.0 kHz. At each BF, parameters were obtained for the DRNL filter to reproduce input/output and tuning curves. The match between the model and the experimental data is almost perfect for frequencies near BF. Quantitatively, the model response gets worse (but is still reasonable) for frequencies well below and well above BF. These discrepancies are discussed in terms of the middle-ear function, which proves critical. The model responses to clicks, AM, multicomponent, and Schroder-phase stimuli were also compared against experimental data. Results show that the architecture of the DRNL filter seems suitable to reproduce this wide range of phenomena. Strategies are discussed for developing a chinchilla nonlinear cochlear filterbank from current parameters. [Work supported by the Consejería de Sanidad of the Junta de Comunidades of Castilla, La Mancha.]

2aPPb14. Estimation of size at three sites on the face. Ronald T. Verrillo (Inst. for Sensory Res., Syracuse Univ., Syracuse, NY 13244), Stanley J. Bolanowski (Inst. for Sensory Res./Dept. of Bioengineering and Neurosci., Syracuse Univ., Syracuse, NY 13244), and Francis P. McGlone (Unilever Res., Bebington, Wirral L63 3J2, UK)

The subjective size of steel balls was judged by nine subjects using the method of Absolute Magnitude Estimation. The balls were rolled (''scripted'') over three areas of the face (forehead, cheek, and lower lip) using the right index distal finger pad. The areas were stimulated under three conditions; (1) the subjects upon themselves (intra-active touch); (2) the subjects upon another person (interactive touch); and (3) another person upon the face of the subject (interactive touch). There were significant differences among the three facial sites in the intra-active condition. When another person's face was scripted by the subject, there were no differences among the sites. The results were mixed when another scripted upon the faces of the subjects. The results are considered in terms of tissue mechanics, mode of stimulation and representation of the sites on the somatosensory cortex. [Work supported by NIH.]

2aPPb15. Underwater loudness for tones. Edward A. Cudahy and Derek Schwaller (Naval Submarine Medical Res. Lab., Groton, CT 06349-5900)

The loudness for pure tones was measured by loudness matching for 1-s pure tones from 100 to 50 000 Hz. The standard tone was 1000 Hz. Subjects were instructed to match the loudness of the comparison tone at one of the test frequencies to the loudness of the standard tone. The standard was presented at one of five sound pressure levels (SPL) for each set of frequencies. The SPL was varied randomly across a test series. The subjects were bareheaded U.S. Navy divers tested at a depth of 3 m. All subjects had normal hearing. The tones were presented to the right side of the subject from an array of underwater sound projectors. The SPL was calibrated at the location of the subject's head with the subject absent. The loudness increased more rapidly as a function of standard SPL at mid-frequencies than at either high or low frequencies. The most compact loudness contours (the least SPL change across the range of standard SPL) were at the highest frequency. Loudness contours across frequency derived

from these measurements are significantly different from in-air measurements with minimum audibility in the 1000 Hz region rather than the 2–4 kHz region observed for in-air measurements.

2aPPb16. Sample discrimination of frequency differences with irrelevant context–revisited. Donna L. Neff, Eric C. Odgaard, and Walt Jesteadt (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

This experiment replicated earlier work to provide additional data for post hoc analyses of perceptual weights for conditions with large effects of context on performance. The non-adaptive, 2AFC task was sample discrimination of frequency differences (SD-F) in quiet and with added context stimuli. Listeners were to select the interval in which a pair of target tones was drawn from the higher of two Gaussian frequency distributions (means of 2000 and 2150 Hz). Pairs of context stimuli were added at frequency regions above and below the targets, at distances of 600, 1000 and 1400 Hz. Context stimuli were fixed-frequency tones, noise bands, or random-frequency tones. All stimuli were 100 ms with 5-ms ramps, presented simultaneously. Conditions were tested with and without Gaussian level jitter (standard deviation=3 dB). Mean levels of target and context stimuli were equated or systematically varied across conditions. The results confirm earlier work, with little effect of fixed-frequency tones or noise-band context at any distance from the targets. Random-frequency context tones produced large detrimental effects even at remote distances, due predominantly to effects of the lower-frequency context tone. Level variation had little effect. Different approaches to calculating perceptual weights influences data interpretation. [Work supported by NIDCD.]

2aPPb17. Estimating the form of the peripheral nonlinearity from slopes of psychometric functions in forward masking. Walt Jesteadt, Stephen T. Neely, and Kim S. Schairer (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131, jesteadt@boystown.org)

In previous work, we have shown that slopes of psychometric functions (PF) obtained for the detection of a signal in the presence of a forward masker are in general agreement with standard assumptions concerning the form of the peripheral nonlinearity. PFs are steep for conditions where signal levels are low and grow progressively shallower as the signal threshold increases in level. PFs become more parallel if signal levels are adjusted to reflect the effect of a nonlinearity of the form proposed by Yates *et al.* [G. K. Yates, I. M. Winter, and D. Robertson, Hear. Res. **45**, 203–220 (1990)]. An alternative approach is to use PF data to estimate the form of the nonlinearity. The data suggest that the reciprocal of the slope of the PF is linearly related to the signal level. This would indicate that the nonlinearity is logarithmic. Use of a logarithmic nonlinearity as the first stage in a model of forward masking predicts thresholds as a function of the masker level and signal delay that are consistent with a formula proposed by Jesteadt *et al.* [W. Jesteadt, S. P. Bacon, and J. R. Lehman, J. Acoust. Soc. Am. **71**, 950–962 (1982)] and predicts loudness growth rates consistent with the literature. [Work supported by NIDCD.]

2aPPb18. Filter asymmetry for short-duration signals as a function of signal delay from masker onset. Veronica Pimentel and Elizabeth A. Strickland (Dept. of Audiol. and Speech Sci., Purdue Univ., West Lafayette, IN 47907, estrick@purdue.edu)

In a previous study, filter shapes were measured for short-duration tones at the onset or in the center of a longer-duration, notched-noise masker. The filter shapes sharpened with signal delay, with the amount of sharpening depending on the level of the signal. The pattern of results was interpreted in terms of a model incorporating published physiological data on nonlinear basilar-membrane input–output functions. The results suggested that these functions might become more linear as the signal is delayed from masker onset. The present study extends this approach to asymmetric notches. This was motivated by previous research suggesting that frequencies above the signal frequency might produce more of a temporal effect than those below the signal frequency, and by physiological data showing that basilar-membrane input–output functions differ for tones above and below the characteristic frequency. The signal was a 10-ms, 4-kHz sinusoid, fixed at levels from 40 to 70 dB SPL, and presented at the onset or in the center of a notched-noise masker. Notches were placed symmetrically or asymmetrically about the signal. The results will be discussed in terms of basilar-membrane input–output functions.

2aPPb19. Similar improvements in auditory discrimination produced by multiple-hour training with a labeling or discrimination procedure. Julia A. Mossbridge and Beverly Wright (Dept. of Commun. Sci. and Disord. and Northwestern Univ. Inst. for Neurosci., 2299 N. Campus Dr., Northwestern Univ., Evanston, IL 60208)

Does the procedure used to train an auditory skill impact the characteristics of subsequent skill improvement? Here, the effects of multiple-hour training with either a labeling or discrimination procedure were compared by using a discrimination procedure to examine performance on six conditions. Listeners completed a pretest on the six conditions, trained on one condition for 720 trials/day for 8 days with either a labeling ($n=6$) or discrimination ($n=6$) procedure, then completed a post-test identical to the pretest. The trained condition for labeling-trained listeners was to determine, for each individual presentation, whether the onsets of two tones were synchronous or asynchronous. The trained condition for discrimination-trained listeners was to determine which of two presentations of these same tones had asynchronous onsets. All listeners showed significant improvements on the trained condition. Further, there were no significant differences between training groups in the pattern of improvement across the six conditions, both immediately and one month after training. Thus, multiple-hour labeling and discrimination training yielded similar learning, suggesting that, in this case, both procedures tapped the same neural substrate. These results inform our understanding of the mechanisms of learning and may guide clinical treatment for those with auditory perceptual disorders. [Work supported by NIDCD.]

2aPPb20. Cut-back frozen noise characterization of auditory sensory memory. Anthony Boemio and David Poeppel (Cognit. Neurosci. of Lang. Lab., 3416 Marie Mount Hall, Univ. of Maryland, College Park, MD 20742, aboemio@glue.umd.edu)

Frozen noise stimuli were comprised of repeated concatenated segments of random noise, with the segment duration determining the periodicity or stimulus onset asynchrony (SOA). Seven conditions were constructed with SOAs ranging from 100–1000 ms. For each of the seven conditions, a fraction of the repeated segments were ''cut-back'' with nonrepeated noise. Subjects were presented with either a frozen noise stimulus or a nonrepeated noise stimulus of equal duration, and told to indicate if they detected periodicity. The thresholds for this two-alternative forced-choice infrapitch periodicity detection task were obtained by an adaptive tracking procedure. The results show that classification performance varies linearly with SOA over the entire 100–1000 ms. range. When normalized by SOA, it becomes evident that detection is characterized by a signal-to-noise ratio (SNR) where the repeated noise is the signal and the nonrepeated noise, the noise. The SNR for all seven conditions was approximately 0.7. It is proposed that cut-back frozen noise stimuli may be used to establish the upper storage limit of auditory sensory memory by extending this technique to SOAs greater than 1000 ms and looking for deviation from linearity in the plot of repeated noise segment vs SOA. [Work supported by NIDCD Grant No. DC 0463801.]

2aPPb21. Perception of duration of ramped and damped sounds with exponential and raised cosine ramps. Massimo Grassi (Dept. of Gen. Psych., Univ. of Padova, via Venezia 8, 35131 Padova, Italy, grassi@psy.unipd.it) and Christopher J. Darwin (Univ. of Sussex, Brighton BN1 9QG, UK)

In the first experiment listeners were asked to adjust the duration of a steady-level sound in order to match the duration of 250-, 500-, 750-, or 1000-ms targets. Targets could be either steady in level, increasing in level with either a raised-cosine or an exponential ramp (ramped sounds), or the ramped sounds reversed in time (damped sounds). The results extended to longer durations than those reported by Schlauch *et al.* [J. Acoust. Soc. Am. **109**, 2880–2887 (2001)]: damped sounds were matched as much shorter (~35% for exponential ramps, ~30% for raised-cosine ramps) and ramped sounds as slightly shorter (~10%) than steady-level sounds. In the second experiment, the subjective duration of all targets used in the first experiment was measured as the difference between reaction times to their onsets and their offsets. Although delays in reaction times to onsets of ramped targets could explain their slight underestimation, anticipations to offsets of damped targets were too small to explain their underestimation. The results of the research contradict the predictions of the AIM model [R. D. Patterson and T. Irino, J. Acoust. Soc. Am. **104**, 2967–2979 (1998)], where ramped and damped sounds longer than 200 ms should be perceived as possessing similar subjective durations.

TUESDAY MORNING, 4 JUNE 2002 KINGS GARDEN NORTH, 8:45 A.M. TO 12:00 NOON

2a TUE. AM

Session 2aSA

Structural Acoustics and Vibration: Pyroshock I

Harry Himelblau, Chair
Space and Communication Division, The Boeing Company, Rocketdyne Propulsion and Power, Canoga Park, California 91309-7922

Invited Papers

8:45

2aSA1. Pyrotechnic devices and their applications. Harry Himelblau (MC IB-31, Rocketdyne Propulsion and Power, The Boeing Co., Canoga Park, CA 91309-7922, harry.himelblau@boeing.com)

Pyroshock is mechanical shock transmitted through structures from explosive devices, sometimes accompanied by structural impact. These devices are designed to cause the intentional separation of structures, or to cause the deployment of various mechanisms or subsystems required for mission operation. Separation devices usually fall into two categories: (a) line sources, such as linear shaped charges, and (b) point sources, such as explosive bolts, pin puller and pushers, and gas generators. The advantages of these devices are high reliability (especially when redundantly activated), low cost and weight, high activation speed, and low structural deformation a short distance from the source. The major limitation is pyroshock, a severe high-frequency transient capable of causing failure or malfunction to small nearby elements, especially electronic and optical components located close to the source. This pyroshock tutorial, which is intended to summarize recent improvements to the technology, is initiated with a review of explosive and companion devices.

9:15

2aSA2. Pyroshock prediction procedures. Allan G. Piersol (Piersol Eng. Co., 23021 Brenford St., Woodland Hills, CA 91364-4830, apiersol@pacbell.net)

Given sufficient effort, pyroshock loads can be predicted by direct analytical procedures using Hydrocodes that analytically model the details of the pyrotechnic explosion and its interaction with adjacent structures, including nonlinear effects. However, it is more common to predict pyroshock environments using empirical procedures based upon extensive studies of past pyroshock data. Various empirical pyroshock prediction procedures are discussed, including those developed by the Jet Propulsion Laboratory, Lockheed-Martin, and Boeing.

9:45

2aSA3. Pyroshock data acquisition—historical developments using piezoelectric accelerometers and other transducers. Harry Himelblau (MC IB-31, Rocketdyne Propulsion and Power, The Boeing Co., Canoga Park, CA 91309-7922, harry.himelblau@boeing.com)

For nearly 50 years, P/E accelerometers have been used for acquiring pyroshock data with mixed results. For longer distances between the explosive source and the transducer location (e.g., two feet or more), valid data of lesser shock magnitude were usually obtained. However, for shorter distances, a variety of problems were often encountered, causing erroneous results. It was subsequentially determined that most problems were caused by measurement system nonlinearities, i.e., the nonlinear resonant response of the

accelerometer, or exceeding the linear amplitude range of the signal conditioner and recorder. In the earlier years, it was erroneously assumed that subsequent low pass filtering of the signal would remove the nonlinearities, hopefully leading to valid data. This only masked the invalid results. Eventually, improved P/E accelerometers were developed with higher natural frequencies and larger amplitude limits that caused substantially fewer problems and allowed measurements closer to the explosive sources. Shortly thereafter, the high frequency noncontact laser doppler vibrometer became available which circumvented the accelerometer resonance problem. However, this velocity transducer is almost always limited to laboratory tests in order to constrain the motion of the laser head by a very rigid and massive support foundation compared to the flexible structure which is attached to the laser target. Other LDV measurement problems have been encountered that must be avoided to achieve valid data. Conventional strain gages have been successfully used to measure pyroshock strain. However, due to the short wavelength of direct and bending pyroshock waves at high frequencies, small strain gages are usually required to avoid spatial averaging over the length of the gage.

10:15–10:30 Break

10:30

2aSA4. Pyroshock data acquisition—recent developments using P/R and P/E accelerometers and isolators. Vesta I. Bateman (Sandia Natl. Labs., P.O. Box 5800, Albuquerque, NM 87185-0553, vibatem@sandia.gov)

Mechanical isolators have been developed for piezoresistive and piezoelectric accelerometers to mitigate high frequency shocks before they reach the accelerometer because the high frequency pyroshocks may cause the accelerometer to resonate and/or break. Several commercial mechanically isolated accelerometers are available to the general public and their characteristics have been studied using Hopkinson bar test techniques. The in-axis response of these devices will be compared. Cross-axis response will be presented for one device. Additionally, pyroshock and ballistic shock measurements, performed by international organizations, will be presented for several isolators.

11:00

2aSA5. Pyroshock data analysis—Analysis, editing, and validation. David O. Smallwood (Sandia National Labs., P.O. Box 5800, Albuquerque, NM 87185-0553, dosmall@sandia.gov)

The most common form of pyroshock data analysis is the shock response spectrum (SRS), which is defined and briefly discussed. The most common way to compute the SRS of a time history is with a digital recursive filter that simulates the response of a single-degree-of-freedom (SDOF) system to the pyroshock excitation. This filter is described together with its strengths and limitations. Sometimes additional characteristics of the time history are used to supplement the SRS. The most common is a definition of duration. If the time history is assumed to be a sum of exponentially decaying sinusoids, a decay rate can be used. Often, the time duration is defined as the duration until the waveform decays to 10% of its peak value. More recently temporal moments have been used to define the temporal characteristics of the waveforms. Flaws in pyroshock data are common, including accelerometer zero shifts, clipped data, dropouts, and instrumentation overloads. Data validation procedures that expose many of these flaws are outlined. Good pyroshock data are difficult to acquire. As a general rule, flawed data should not be used, but sometimes we are forced to use flawed data. Procedures are briefly described for editing flawed data that will result in credible, although not necessarily correct, results.

11:30

2aSA6. Pyroshock data analysis—The GCPP validation procedure. Allan G. Piersol (Piersol Eng. Co., 23021 Brenford St., Woodland Hills, CA 91364-4830, apiersol@pacbell.net)

This procedure was developed for validating pyroshock data by Powers and the author, using techniques originated by Gaberson and Chalmers in validating naval shock data at lower frequencies and subsequently modified. It requires that the acceleration time history be single and double integrated to obtain a velocity and a displacement time history, which is to be examined. Valid data look like a low-pass filtered acceleration time history which is not integrated. The maximum resulting displacement should approximate the independently measured displacement (often near zero) or that previously computed analytically for the structure at the measured location. In addition, the positive and negative SRS are computed and compared. Valid data usually show similar spectral content.

Session 2aSC

Speech Communication: Cross-Linguistic Studies: Acoustics and Perception (Poster Session)

Kenneth de Jong, Chair
Department of Linguistics, Indiana University, 322 Memorial Hall, Bloomington, Indiana 47403

Contributed Papers

All posters will be on display from 8:30 a.m. to 5:00 p.m.. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 12:00 noon. To allow for extended viewing time, posters will remain on display until 5:00 p.m.

2aSC1. The acquisition of *L*2 phonetic categories: Single-, two-, and multiple-category assimilation in Dutch learners of Spanish. Paola Escudero (Utrecht Univ./McGill Univ., Trans 10, 3512 JK Utrecht, The Netherlands, paola.escudero@let.uu.nl) and Paul Boersma (Univ. of Amsterdam, The Netherlands)

The perceptual acquisition of Spanish vowel categories by Dutch speakers is examined. One might think that these learners have few problems because Dutch has as many as 12 monophthongs in its vowel inventory, and Spanish has only 5. However, a perception experiment shows that Dutch learners make quite a few mistakes when identifying Spanish vowels. It was predicted that these errors exemplify three well-attested patterns in the initial state of *L*2 acquisition: single-, two-, and multiple-category assimilation. These patterns may be found in the categorization of Spanish /i/–/e/, /e/–/a/, and /u/–/o/, respectively. The strategies that the learners follow to solve the problems associated with these patterns are presented, as well as the development of phonetic categorization in the three scenarios. Ninety-six listeners were tested with a three-way experiment design. The stimuli were Spanish CVCs embedded in Dutch or Spanish carrier sentences. The first two experiments were designed to find out to what degree learners perceive *L*2 sounds differently from *L*1 sounds. The third experiment tested the development of the learners' *L*2 categorization accuracy. The combined results of the three experiments show that the learners indeed manifest the three different assimilation patterns, each of which is solved by a different strategy.

2aSC2. Phonetic discrimination and non-native spoken-word recognition. Andrea Weber (CUNY Grad. Ctr., Speech and Hearing Sci., 365 Fifth Ave., New York, NY 10016, aweber@gc.cuny.edu) and Anne Cutler (Max Planck Inst. for Psycholinguist., Wundtlaan 1, 6525 XD Nijmegen, The Netherlands, anne.cutler@mpi.nl)

When phoneme categories of a non-native language do not correspond to those of the native language, non-native categories may be inaccurately perceived. This may impair non-native spoken-word recognition. Weber and Cutler investigated the effect of phonetic discrimination difficulties on competitor activation in non-native listening. They tested whether Dutch listeners use English phonetic contrasts to resolve potential competition. Eye movements of Dutch participants were monitored as they followed spoken English instructions to click on pictures of objects. A target picture (e.g., picture of a paddle) was always presented along with distractor pictures. The name of a distractor picture either shared initial segments with the name of the target picture (e.g., target paddle, /paedl/ and competitor pedal, /pEdl/) or not (e.g., strawberry and duck). Half of the target-competitor pairs contained English vowels that are often confused by Dutch listeners (e.g., /ae/ and /E/ as in ''paddle-pedal''), half contained vowels that are unlikely to be confused (e.g., /ae/ and /aI/ as in ''parrot-pirate''). Dutch listeners fixated distractor pictures with confusable English vowels longer than distractor pictures with distinct vowels. The results demonstrate that the sensitivity of non-native listeners to phonetic contrasts can result in spurious competitors that should not be activated for native listeners.

2aSC3. Perception of the English intrusive stops by Korean listeners. Jeong-Im Han (Dept. of English, Konkuk Univ., Hwayang-dong 1, Gwangjin-gu, Seoul 143-701, Korea, jhan@konkuk.ac.kr)

This paper reports results of an experiment examining Korean listeners' perception of English intrusive stops in nasal-obstruent clusters. The experiment tests (1) how often intrusive stops are perceived; (2) how language-specific syllable structure constraints influence the perception, given the fact that Korean does not allow consonant clusters in syllable onsets and codas; (3) whether even the perception of phonetic variables like intrusive stops, not phonemes, could be improved by learning. Ninety English non-words with a monosyllable structure of CVC1C2 were created, where C1=/m,n,N/, and C2=/p,k,s/. The stimuli including additional 90 filler items were recorded by three native English speakers and one representative data among them was given to three groups of native Korean listeners in terms of their English proficiency. Each was asked to monitor the target sounds [Warner and Weber, J. Phonetics **29**, 23–52 (2001)]. The preliminary results show that identification of intrusive stops in English is totally dependent on Korean syllable structure, so even stimuli with strong acoustic cues were misparsed. Nonetheless, there's a high correlation between perception of intrusive stops and listeners' English proficiency, showing the possibility of the improvement of perception by learning.

2aSC4. Effects of consonantal context on perception of French rounded vowels by American English adults with and without French language experience. Erika S. Levy and Winifred Strange (Speech and Hearing Sci., City Univ. of New York, Grad. School and Univ. Ctr., 365 5th Ave., New York, NY 10016, erilev@earthlink.net)

This study extended Gottfried's [J. Phonetics **12**, 91–114 (1984)] investigation of the effects of learning French on AE listeners categorial discrimination of the contrasts involving Parisian French vowels /y/, /ø/, /u/, and /i/. Vowels were presented in /rabVp/ and /radVt/ bisyllables embedded in carrier phrases by three different speakers in an AXB discrimination task. Two groups were tested: proficient *L*2-French-speaking AE listeners (Exp) and non-French-speaking AE listeners (Inexp). Overall, the Exp group performed better than the Inexp group on /u-ø/, /y-i/, and /y-ø/ distinctions (mean errors: Exp=5%, Inexp=24%). However, for /u-y/, the groups did not differ (Exp=30% vs Inexp=24% errors). The findings may be explained by perceptual assimilation patterns, which are determined, in part, by consonantal context. The Inexp group confused /y/ with /i/ in bilabial context, but /y/ with /u/ in alveolar context, whereas the Exp group

confused /y/ with /u/, in both contexts. For all contrasts, the Inexp group performed better in bilabial than in alveolar context (16% vs 32% errors), whereas no context effect was revealed for the Exp group. The results suggest that learning a foreign language includes learning its coarticulatory rules.

2aSC5. Perception of English final stops by native Spanish and Russian listeners. Paul Argodale and Winifred Strange (CUNY Grad. Ctr., 365 Fifth Ave., New York, NY 10016)

Spanish and Russian speakers perception of word-final stops was investigated in an identification task. Natural stimulus triads (sue-suit-sued, see-seat seed) were produced and presented in two conditions: (1) Final consonants released and CV in the utterance final position; (2) final consonants unreleased and CV produced before a word beginning with a voiceless stop. Russian listeners, whose first language has final voiceless stops, correctly identified English final stops better than Spanish listeners, whose first language does not have final stops (72% vs 60% correct overall). Spanish listeners were more likely than Russian listeners to identify words with final stops as open syllables (31% vs 2% errors). A significant effect of release condition was also found: Spanish listeners were much better at identifying released stops than unreleased stops (90% vs 29% correct). Russian listeners also showed a slightly higher correct rate of identification for released stops (91% vs 42%). Both groups identified voiceless final stops more accurately than voiced stops: Russians (80% vs 65% correct); Spanish (65% vs 54% correct). These results replicate and extend earlier work by Flege and colleagues. [Work supported by a grant from the Professional Staff Congress of the City University of New York.]

2aSC6. Interlanguage intelligibility benefit as a function of talker and listener language background and *L*2 proficiency. Tessa Bent and Ann R. Bradlow (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, t-bent@northwestern.edu)

For non-native listeners, intelligibility of non-native speakers (NNS) can surpass intelligibility of native speakers [T. Bent, J. Acoust. Soc. Am. **109**, 2472 (2001)]. The present study further investigated how language background and *L*2 proficiency of talker and listener affect *L*2 intelligibility. Two Chinese NNS, two Korean NNS, and one English monolingual were recorded reading simple English sentences. Listeners were English monolinguals, Chinese NNS, Korean NNS, and NNS with other *L*1s. For native English listeners, the native English talker was most intelligible. For non-native listeners, intelligibility of non-native talkers with high degrees of *L*2 proficiency was better than or equal to native talker intelligibility. This pattern held regardless of whether non-native talkers and listeners matched in *L*1. Additionally, the interlanguage benefit (i.e., the intelligibility difference between non-native and native talkers) decreased as the listeners' proficiency in English increased. This interlanguage benefit can be accounted for by a combination of a shared systematic interlanguage (when talker and listener match in *L*1) and the influence of interlanguage universals (when talker and listener do not match in *L*1). However, both of these factors become less influential as listeners *L*2 proficiency develops causing a decrease in the interlanguage benefit. [Work supported by NIH-NIDCD Grant DC 03762.]

2aSC7. Phonological awareness of English by Chinese and Korean bilinguals. Hyunjoo Chung, Anna Schmidt, and Tse-Hsuan Cheng (School of Speech Pathol. and Audiol., Kent State Univ., Kent, OH 44242, aschmidt@kent.edu)

This study examined non-native speakers phonological awareness of spoken English. Chinese speaking adults, Korean speaking adults, and English speaking adults were tested. The L2 speakers had been in the US for less than 6 months. Chinese and Korean allow no consonant clusters and have limited numbers of consonants allowable in syllable final position, whereas English allows a variety of clusters and various consonants in syllable final position. Subjects participated in eight phonological awareness tasks (4 replacement tasks and 4 deletion tasks) based on English phonology. In addition, digit span was measured. Preliminary analysis indicates that Chinese and Korean speaker errors appear to reflect L1 influences (such as orthography, phonotactic constraints, and phonology). All three groups of speakers showed more difficulty with manipulation of rime than onset, especially with postvocalic nasals. Results will be discussed in terms of syllable structure, L1 influence, and association with short term memory.

2aSC8. Interaction of cues to vowel identity and consonant voicing: Cross-language perception. Geoffrey Stewart Morrison (Dept. of Linguist., Simon Fraser Univ., Burnaby, BC V5A 1S6, Canada)

Canadian English has two high front vowels differing in spectral and duration properties, Spanish has one high front vowel, and Japanese has two high front vowels differing in duration only. Vowel duration is a major cue to post-vocalic consonant voicing in English, but not in Japanese or Spanish. Canadian English, Japanese, and Mexican Spanish listeners identified members of a multidimensional edited speech continuum covering the English words *bit, beat, bid, bead*. The continuum was created by systematically varying the spectral properties of the vowel, and the durations of the vowel, the consonant closure, and the carrier sentence. English listeners had a categorical cutoff between /i/ and /ɪ/ based primarily on the spectral properties of the vowel. Half the English listeners identified consonant voicing using vowel duration. Japanese listeners had a categorical cutoff between the English vowels based primarily on the duration of the vowel. The location of the cutoff was the same as the categorical cutoff between Japanese long /i:/ and short /i/. Japanese listeners identified consonant voicing at random. Spanish listeners identified the English vowels using vowel duration but did not have a categorical cutoff. Half the Spanish listeners identified consonant voicing using the spectral properties of the vowel.

2aSC9. Perception of English resyllabification by monolingual Japanese listeners. Kyoko Nagao, Byung-jin Lim, and Kenneth de Jong (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405)

Previous results (Nagao *et al.,* 2001) show that non-native listeners exhibit remarkably similar patterns of perceptual resyllabification with English listeners, suggesting that perceptual resyllabification is not a language specific phenomenon. The same listeners tended to identify voiced English tokens as voiceless, in keeping with Japanese voicing categories. In order to determine the degree to which the non-native perceptual resyllabification was due to extensive exposure to English, monolingual Japanese listeners, 8 from an older and 12 from a younger generation, participated in the same experiment. Monolinguals showed perceptual resyllabification of the same tokens as do English listeners', consistent with the previous results. Also consistent with previous results, especially older listeners' responses were more affected by Japanese voicing categories. In addition, older listeners were more likely than English listeners to identify tokens as CV's, while younger listeners were less likely to. The results for older listeners are what is expected of biases toward native categories. The results for the younger listeners, however, seem to indicate a developmental process involving the construction of new prosodic categories. [Work supported by NIDCD and NSF.]

2aSC10. Cross-language perception of fricative/vowel syllables by speakers of English and Mandarin. Sreedivya Radhakrishnan and Anna Schmidt (School of Speech Pathol. & Audiol., P.O. Box 5190, Kent State Univ., Kent, OH 44242, sradhakr@kent.edu)

A recent study [Radhakrishnan and Schmidt (2001)] on the perception of fricative/vowel syllables by English and Mandarin speakers demonstrated that English speakers use the fricative noise while Mandarin speakers use both the fricative noise and vowel transitions in categorizing syl-

lable initial fricatives in their native languages. In this study, cross-language mappings for fricative/vowel syllables were obtained from native English and Mandarin speakers. Native English and Mandarin speakers labeled and rated three sets of tokens from L2. These tokens were (1) fricative noises from four different vowel environments, (2) natural fricative vowel syllables, and (3) F+V combinations in which the fricative noise from one language was combined with vowel transition from the other language. Vowel length was manipulated in the third set of stimuli. Results indicate differences in the cross-language mappings by the two language groups.

2aSC11. Learning to perceive English /d/ versus /ð/: A comparison of French, English, and English–French bilingual 4-year-olds. Megha Sundara, Linda Polka (School of Commun. Sci. and Disord., McGill Univ., 1266 Pine Ave. W., Montreal, QC H3G 1A8, Canada), Lisa Campisi, Fred Genesee, and Caroline Marcoux (McGill Univ., Montreal, QC H3A 1B1, Canada)

Recent findings show that discrimination of the English /d–ð/ does not differ for English and French infants (6–8-month-olds and 10–12-month-olds), although English adults clearly outperform French adults on this contrast, which is not phonemic in French. With respect to age effects, English listeners' perception of /d–ð/ improves between infancy and adulthood, whereas French listeners' perception remains unchanged [Polka *et al.*, J. Acoust. Soc. Am. **109**, 2190–2200 (2001)]. In the present study, we tested monolingual English, monolingual French, and early English–French bilingual 4-year-olds on the same contrast using the same stimuli and procedures to clarify when facilitative effects of language experience emerge and whether they are affected by bilingualism. Four findings are reported. First, a language effect (English>French) is evident by 4 years of age. Second, among native (English) listeners facilitative effects are evident by 4-years of age (infants<4-year-olds<adults). Third, among non-native (French) listeners discrimination performance is comparable across the age groups tested (infants=4-year-olds=adults). Fourth, bilingual 4-year-olds' performance is virtually identical to that of their French-speaking peers, revealing a strong effect of bilingualism on the perception of this contrast. Several factors contributing to these findings will be discussed.

2aSC12. On the role of phonetic inventory in the perception of foreign-accented speech. Joan Sereno, Joyce McCall, Allard Jongman (Linguist. Dept., The Univ. of Kansas, 1541 Lilac Ln., Lawrence, KS 66044-3177, sereno@ku.edu), Ton Dijkstra, and Walter van Heuven (Univ. of Nijmegen, 6500 HE Nijmegen, The Netherlands)

The current study investigates the effect of phonetic inventory on perception of foreign-accented speech. The perception of native English speech was compared to the perception of foreign-accented English (Dutch-accented English), with selection of stimuli determined on the basis of phonetic inventory. Half of the stimuli contained phonemes that are unique to English and do not occur in Dutch (e.g., [θ] and [æ]), and the other half contained only phonemes that are similar in both English and Dutch (e.g., [s], [i]). Both word and nonword stimuli were included to investigate the role of lexical status. A native speaker of English and a native speaker of Dutch recorded all stimuli. Stimuli were then presented to 40 American listeners using a randomized blocked design in a lexical decision experiment. Results reveal an interaction between speaker (native English versus native Dutch) and phonetic inventory (unique versus common phonemes). Specifically, Dutch-accented stimuli with common phonemes were recognized faster and more accurately than Dutch-accented stimuli with unique phonemes. Results will be discussed in terms of the influence of foreign accent on word recognition processes.

2aSC13. Spectral resolution and English experience: effects on English phoneme and word recognition by non-native English speakers. Monica Padilla (Dept. of Biomed. Eng., Univ. of Southern California, Los Angeles, CA) and Robert V. Shannon (House Ear Inst., Los Angeles, CA)

Normal hearing listeners whose first language was Spanish were tested with English phonemes, words and sentences. Listeners were divided into four categories according to experience with the second language. Speech was presented in a sound treated booth at a level of 70 dBA. Listening conditions included noise (SNR of 15 dB, 10 dB, 5 dB, 0 dB, and 5 dB) and reduced spectral information (2, 4, 6, 8 and 16 frequency bands). Plomp's Model [J. Speech Hear. Res. **29**, 146–154 (1986)] was applied to the data. The distortion factor "D" defined by Plomp was found to increase with an increased loss of spectral resolution. It was also found to increase with age of learning of the second language. An additional "distortion" seems to be introduced when a second language is learned at a later age. Non-native listeners had more difficulty understanding vowels, words and sentences. Surprisingly, English experience had less effect on word and sentence recognition than on vowel recognition. Significantly lower performance on vowel recognition was seen even for fully bilingual listeners with reduced spectral resolution which could probably be related to the conflicting vowel spaces of the two languages. [Work funded by NIDCD.]

2aSC14. Language-specific effects in speaker and rate normalization. Allard Jongman, Yuwen Lai (Linguist. Dept., The Univ. of Kansas, 1541 Lilac Ln., Lawrence, KS 66044-3177, jongman@ku.edu), and Corinne Moore (ProQuest, Richfield, OH 44286)

This study concerns the effects of context on the perception of Mandarin tones. Mandarin tones are primarily distinguished in terms of $F0$ range and temporal location of inflection point. The most relevant context effects therefore involve changes in speaker and speaking rate. In this study, we compared speaker and rate normalization of Mandarin tones 2 (mid–rising) and 3 (low–falling–rising) by Mandarin and English listeners. Synthetic tone 2–tone 3 continua varying in turning point (TP, the inflection point of the tone), $\Delta F0$ (the difference in $F0$ between onset and turning point), or both were appended to natural precursor phrases representing high- and low-pitched speakers, as well as fast and slow speaking rates. Results of the rate experiments with Mandarin listeners confirmed that normalization occurs only when precursors and stimuli vary along the same acoustic dimension, a finding previously observed for speaker normalization [C. B. Moore and A. Jongman, J. Acoust. Soc. Am. **102**, 1864–1877 (1997)]. English speakers showed speaker and rate normalization effects as well, but only for the continuum in which TP and $\Delta F0$ varied simultaneously. We propose that limitations on perceptual resources allow English listeners to attend to extrinsic information only when intrinsic acoustic differences become more perceptually salient.

2aSC15. Effects of training on learning non-native speech contrasts. Joan M. Sinnott (Psych. Dept., Univ. of South Alabama, Mobile, AL 36688)

An animal psychoacoustic procedure was used to train human listeners to categorize two non-native phonemic distinctions. In Exp 1, Japanese perception of the English liquid contrast /r-l/ was examined. In Exp 2, American-English perception of the Hindi dental-retroflex contrast /d-D/ was examined. The training methods were identical in the two studies. The stimuli consisted of 64 CVs produced by four different native talkers (two male, two female) using four different vowels. The procedure involved manually moving a lever to make either a "go-left" or "go-right" response to categorize the stimuli. Feedback was given for correct and incorrect responses after each trial. After 32 training sessions, lasting about 8 weeks, performance was analyzed using both percent correct and response time as measures. Results showed that the Japanese listeners, as a group, were statistically similar to a group of native listeners in categorizing the liquid contrast. In contrast, the Amercan-English listeners were not

nativelike in categorizing the dental-retroflex contrast. Hypotheses for the different results in the two experiments are discussed, including possible subject-related variables. In addition, the use of an animal model is proposed to objectively ''calibrate'' the psychoacoustic salience of various phoneme contrasts used in human speech.

2aSC16. Listeners' perceptions of and attitudes toward Hispanic-, Asian-, and Arabic-accented English speakers. Norman Lass, Traci Atkins (Dept. of Speech Pathol. & Audiol., West Virginia Univ., Morgantown, WV 26506-6122), and Rebekah Squires (James Madison Univ., Harrisonburg, VA 22807)

A master tape containing the randomized recordings of 20 Hispanic-, Asian-, and Arabic-accented English speakers reading a standard prose passage was presented to a group of 22 native English-speaking listeners who participated in two listening sessions. In the first session they were asked to use a 5-point listening preference rating scale. In the second session they heard the same speakers and were asked to determine the presence or absence of an accent and, if present, the type (Asian, Hispanic, Arabic, or other) and degree (mild, moderate, or severe) of accentedness. A scattergram plotting listeners' mean listening preference ratings and degree of accentedness ratings for each of the speakers in the study revealed a strong inverse relationship which yielded a statistically significant ($p < 0.01$) correlation coefficient.Thus, the higher the degree of severity of the listeners' judged accentedness of speakers, the more negative their listening preference rating judgments of the speakers. Implications of these findings and suggestions for future research are discussed.

2aSC17. An acoustical study of English word stress produced by Americans and Koreans. Byunggon Yang (English Dept., Dongeui Univ., 24 Kayadong, Pusanjingu, Pusan 614-714, South Korea)

Acoustical correlates of stress can be divided into duration, intensity, and fundamental frequency. This study examined the acoustical difference in the first two syllables of stressed English words produced by ten American and Korean speakers. The Korean subjects scored very high in TOEFL. They read, at a normal speed, a fable from which the acoustical parameters of eight words were analyzed. In order to make the data comparison meaningful, each parameter was collected at 100 dynamic time points proportional to the total duration of the two syllables. Then, the ratio of the parameter sum of the first rime to that of the second rime was calculated to determine the relative prominence of the syllables. Results showed that the durations of the first two syllables were almost comparable between the Americans and Koreans. However, statistically significant differences showed up in the diphthong pronunciations and in the words with the second syllable stressed. Also, remarkably high r-squared values were found between pairs of the three acoustical parameters, which suggests that either one or a combination of two or more parameters may account for the prominence of a syllable within a word. [Work supported by Korea Science Foundation R01-1999-00229.]

2aSC18. Perception of cross-language vowel differences: A longitudinal study of native Spanish learners of English. Satomi Imai (Div. of Speech and Hearing Sci., Univ. of Alabama at Birmingham, Birmingham, AL 35294, imais@uab.edu), James Flege (Univ. of Alabama at Birmingham, Birmingham, AL 35294), and Ratree Wayland (Univ. of Florida at Gainesville, Gainesville, FL 32611-5454)

The study evaluated the ability of native Spanish speakers to perceive phonetic differences between Spanish vowels (/i e a o u/) and English vowels (/ieɪɑoʊu/). Eighteen adult native speakers of Spanish who were learning English as a second language (L2) in Birmingham. AL were tested at 6-month intervals over a 3.5-year period (T1–T7). Five tokens of each Spanish and English vowel were randomly presented for classification in terms of one of the five vowels of Spanish, and were rated for goodness of fit on a 6-point scale (where 0 indicated ''not Spanish'' and 5 indicated a ''good example'' of a Spanish vowel). At both T1 and T7, English /i/, /eɪ/, /ɑ/, /oʊ/ and /u/ were usually classified as Spanish /i/, /e/, /a/, /o/ and /u/, respectively. A group analysis revealed that significantly higher ratings were given to the Spanish than English member of the /o/-/oʊ/, /u/-/u/, /e/-/eɪ/ and /a/-/ɑ/ pairs but not the /i/-/i/ pair. However, the number of individual L2 learners who gave significantly higher ratings to a Spanish vowel than to the corresponding English vowel differed substantially (/o/-/oʊ/ $n=18$, /u/-/u/ $n=8$, /e/-/eɪ/ $n=5$, /a/−/ɑ/ $n=1$, /i/-/i/ $n=1$). Possible explanations for these between-pair differences —acoustic and perceptual—will be discussed. [Work supported by NIH.]

2aSC19. Release bursts in English word-final stops: A longitudinal study of Korean adults' and children's production. Kimiko Tsukada (Div. of Speech & Hearing Sci., Univ. of Alabama at Birmingham, CH20, 1530 3rd Ave. S., Birmingham, AL 35294, tsukadak@shrp.uab.edu), Molly Mack (Univ. of Illinois, Urbana, IL 61801), Hyekyung Sung (Stanford Univ., Stanford, CA 94305), David Birdsong (Univ. of Texas, Austin, TX 78712), and Ellen Bialystok (York Univ., Toronto, ON M3J 1P3, Canada)

Stops at the end of Korean words are always unreleased. The question addressed here was whether Korean adults and children living in the U.S. can learn to release stops at the end of English words. Four groups of 18 native Koreans (NK) who differed according to age (adult versus child) and length of residence in the U.S. (3 vs 5 years at T1) participated. Two native English (NE) groups served as age-matched controls. Production data were collected at two times (T1, T2) separated by one year. English words ending in /t/ and /k/ were then examined in perception experiments (Exp. 1, Exp. 2). NE-speaking judges decided whether the final stop has a release burst or not. Exp. 1 showed that NE talkers released /t/ more often than NK talkers did. The effect of time was also significant. Talkers produced release bursts more often at T2 than at T1. Exp. 2 showed that, unlike Exp. 1, there were significant differences between NK adults and children. While NK children did not differ from NE children, NK adults released the final /k/ much less often than NE adults did. Possible reasons for why the expected children's advantage was seen for /k/, but not for /t/, will be discussed. [Work supported by NIH.]

2aSC20. An acoustic study of hiatus resolution in two Romance languages. Ioana Chitoran (Linguist. & Cognit. Sci., Dartmouth College, HB 6087, Hanover, NH 03755) and Jose Ignacio Hualde (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

Spanish and Romanian contrast vowel sequences [CiV] in hiatus and corresponding diphthongs [CjV], with some interspeaker variation. Both languages contain surface diphthongs derived historically by gliding, /CiV/ > [CjV]. They both show a strong tendency for blocking gliding word-initially, supported by native speaker judgments: Sp. [miope], Rom. [miopu] short-sighted; Sp. [italjana], Rom. [italjana] Italianf. Data from six speakers of each language confirmed this variation. The duration and $F2$ transition rate of the vocalic sequence were compared, in words containing [i.a] and [ja]. [ia] was significantly longer in hiatus, and had a significantly slower transition rate than in [ja]. However, the ranges of the hiatus and [ja] sets showed some overlap, suggesting that hiatus resolution is not a categorical phonological process. Instead, lexical items fall on a hiatus-to-[ja] continuum. Further comparison of ranges and standard deviations confirmed the correlation between this variation and word position. In both languages more word-initial sequences resist gliding. Significantly less variation was found in the range for hiatus word-initially than for [ja] word-medially. This suggests that the combination of lingual gestures between high and nonhigh vowels is more tightly controlled word-initially than word-medially, a result previously reported for stop sequences [Byrd (1996); Chitoran, Goldstein, and Byrd (unpublished)].

2aSC21. Segmental and coarticulatory timing in Ikalanga prenasalized stops. Patrice S. Beddor, Chutamanee Onsuwan, and Rose Letsholo (Dept. of Linguist., Univ. of Michigan, Ann Arbor, MI 48109, beddor@umich.edu)

The temporal characteristics of prenasalized stops in Ikalanga, a Bantu language of Botswana, are investigated. Eight speakers were recorded producing Ikalanga words with $V^{N}CV$ (^{N}C=/^{m}b ^{n}d $ŋ_{g}$), VNV (N=/mnŋ), and VCV (C=/b d g/) sequences. Acoustic measures of 3 speakers' productions showed that oral closure durations (from vowel offset to closure release) were only slightly longer for ^{N}C than for N or C (94, 83, and 83 ms, respectively). When analyzed separately, the ^{N} and C portions of $V^{N}CV$ were somewhat shorter than their counterparts in VNV and VCV, but nearly all of the shortening occurred in the C portion (in $V^{N}CV$, ^{N}=78 ms and C closure=16 ms). The durations of vowels preceding the medial consonants did not differ across the three consonant types. Thus, for segmental durations, the main difference between Ikalanga $V^{N}CV$ and VNV was the presence of a brief oral occlusion in the former. However, measures of coarticulatory nasalization point to another systematic difference: while both sequences exhibited comparable anticipatory vowel nasalization, VNV (but not $V^{N}CV$) sequences showed carryover nasalization extending through more than 70% of the post-nasal vowel. Findings will be compared with previous work on Bantu prenasalized stops. [Work supported by NSF.]

2aSC22. Factors affecting the duration of native and second-language sentences produced in unspeeded and speeded conditions. James E. Flege (Div. of Speech and Hearing Sci., Univ. of Alabama at Birmingham, Birmingham, AL 35294, jeflege@uab.edu) and Ian R. A. MacKay (Univ. of Ottawa, Ottawa, ON K1N 6N5, Canada)

The aim of this study was to provide insight into why non-natives generally produce longer second-language (L2) sentences than native speakers do. Four groups of 16 Italian–English bilinguals were recruited in Ottawa based on orthogonal differences in age of arrival (AOA) to Canada from Italy (early versus late) and self-reported percentage Italian use (low-L1 use versus high-L1 use). The bilinguals repeated duration-matched English and Italian sentences presented via a loudspeaker in an ''unspeeded'' condition, then as rapidly as possible (the ''speeded'' condition). The same effect of AOA was obtained in both conditions despite a 20% reduction in sentence duration in the speeded condition. That is, the early bilinguals produced significantly shorter English than Italian sentences, whereas the late bilinguals produced significantly longer English than Italian sentences. Different L1 use effects were obtained in the two conditions, however. In the unspeeded condition, the low-L1-use bilinguals produced shorter English than Italian sentences, with a nonsignificant effect of language for high-L1-use bilinguals. In the speeded condition, high-L1-use bilinguals produced shorter Italian than English sentences, with no language effect for the low-L1-use bilinguals. The underlying bases of AOA and L1 use effects on L2 sentence production will be discussed. [Work supported by NIH.]

2aSC23. Duration and *F*0 cues for vowel length in Japanese. Tomoko Kozasa (Dept. of Linguist., Univ. of Hawaii at Manoa, 1890 East–West Rd., Moore Hall 596, Honolulu, HI 96822)

Two acoustic cues for the long–short vowel contrast were investigated in productions of Japanese accented vowels under two rates of speech. An accented word has a high–low pitch sequence. Long accented vowels have the fall within the syllable, while short vowels do not. Unsurprisingly, the vowel duration was significantly shorter in fast speech than in slow speech overall. The speech rate affected the long vowel duration more than the short vowel duration; the ratio between short vowels and long vowels varied by speech rate: 1:1.7 in fast speech and 1:1.9 in slow speech. However, the pitch fall was not significantly affected by the speech rate in either type of vowel. These results suggest that for native Japanese speakers, pitch information may be a more stable cue than duration for distinguishing the length of accented vowels. A perception study manipulating these cues is in progress. Synthesized long vowels without the pitch fall characteristic of long accented vowels are compared with normally accented long and short vowels. If Japanese speakers perceive the synthesized vowels as short, this constitutes evidence that for accented vowels, the pitch cue dominates the duration cue.

2aSC24. Acquisition of Japanese contracted sounds in L1 phonology. Chiharu Tsurutani (School of Lang. and Linguist., Griffith Univ., Nathan 4111 QLD, Australia, C.Tsurutani@mailbox.gu.edu.au)

Japanese possesses a group of palatalized consonants, known to Japanese scholars as the contracted sounds, [CjV]. English learners of Japanese appear to treat them initially as consonant + glide clusters, where there is an equivalent [Cj] cluster in English, or otherwise tend to insert an epenthetic vowel [CVjV]. The acquisition of the Japanese contracted sounds by first language (L1) learners has not been widely studied compared with the consonant clusters in English with which they bear a close phonetic resemblance but have quite a different phonological status. This is a study to investigate the L1 acquisition process of the Japanese contracted sounds (a) in order to observe how the palatalization gesture is acquired in Japanese and (b) to investigate differences in the sound acquisition processes of first and second language (L2) learners: Japanese children compared with English learners. To do this, the productions of Japanese children ranging in age from 2.5 to 3.5 years were transcribed and the pattern of misproduction was observed.

2aSC25. Acoustic and perceptual evidence of complete neutralization of word-final tonal specification in Japanese. Kazumi Maniwa (Dept. of Linguist., Univ. of Kansas, 1541 Lilac Ln., 427 Blake, Lawrence, KS 66044-3177)

This study investigated the extent to which the Japanese lexical pitch-accent distinction is neutralized in word-final position. Native speakers of Tokyo Japanese produced minimal word pairs differing in final accent status. Words were produced both in isolation and in a sentential context, where neutralization would not be expected due to following tonal specification. Examination of pitch patterns on relevant moras revealed a clear distinction between accent-opposed pairs produced in context but no such difference between items produced in isolation. Both the words produced in isolation and the words excised from sentential contexts were then presented to Japanese listeners in a lexical identification task. Participants could clearly distinguish items extracted from sentences but identified words uttered in isolation at chance level. These results suggest that phonological neutralization of final pitch accent is complete, showing no effects of underlying specification in either production or perception.

2aSC26. Perception and production of word-initial long fricatives in Polish. Ela Thurgood (Dept. of English, California State Univ., Chico, CA 95929-0830)

The paper provides an analysis of the perception and production of the Polish distinction between long fricative /v:/ and short fricative /v/ in pre-vocalic position. The task involved 27 native speakers of Polish repeating 50 sentences, including the target sentences. The target sentences were minimal pairs with the meaning difference based solely on fricative length. The results showed that the 190-ms mean duration of the long fricatives was well over twice as long as the 71-ms duration of the short fricatives. When duration of the fricative fell below 147 ms, the listener's ability to discriminate between a short and long fricative was significantly reduced. Fricative length correlated with the duration of the following vowel: 78% of the time, the vowel was longer in the syllable with the long fricative. Indeterminacy about the identification of long fricatives is particularly well marked in these data: 42% of the responses to prompts with long fricatives included a distinctive pitch pattern, a so-called flat hat: a rising F0 in the first syllable, pitch height maintenance, and a drop after the second syllable. In Polish, this pattern indicates that the speaker is paying particular attention to this part of the utterance [Demenko (1999)].

2aSC27. Tone classification of syllable-segmented Thai speech based on multilayer perception. Nuttavudh Satravaha, Powsiri Klinkhachorn (Ln. Dept of Computer Sci. & Elec. Eng., West Virginia Univ., Morgantown, WV 26506), and Norman Lass (West Virginia Univ., Morgantown, WV 26506)

Thai is a monosyllabic tonal language that uses tone to convey lexical information about the meaning of a syllable. Thus to completely recognize a spoken Thai syllable, a speech recognition system not only has to recognize a base syllable but also must correctly identify a tone. Hence, tone classification of Thai speech is an essential part of a Thai speech recognition system. Thai has five distinctive tones (''mid,'' ''low,'' ''falling,'' ''high,'' and ''rising'') and each tone is represented by a single fundamental frequency ($F0$) pattern. However, several factors, including tonal coarticulation, stress, intonation, and speaker variability, affect the $F0$ pattern of a syllable in continuous Thai speech. In this study, an efficient method for tone classification of syllable-segmented Thai speech, which incorporates the effects of tonal coarticulation, stress, and intonation, as well as a method to perform automatic syllable segmentation, were developed. Acoustic parameters were used as the main discriminating parameters. The $F0$ contour of a segmented syllable was normalized by using a z-score transformation before being presented to a tone classifier. The proposed system was evaluated on 920 test utterances spoken by 8 speakers. A recognition rate of 91.36% was achieved by the proposed system.

2aSC28. An acoustical study of the voicing distinction in Dutch plosives. Petra M. van Alphen (MPI for Psycholinguist., P.O. Box 310, 6500 AH Nijmegen, The Netherlands, petra.vanalphen@mpi.nl)

Dutch has two voiced plosives, namely /b/ and /d/. They are said to have a negative VOT (i.e., are prevoiced), while the voiceless plosives /p/ and /t/ have a small, positive VOT. A production study was designed to investigate two questions. First, how does prevoicing vary in spoken Dutch? Second, what other cues in the acoustic signal might contribute to the perceptual distinction between voiced and voiceless plosives? Ten subjects were asked to produce a list of monosyllabic words which were chosen to study the following factors: place of articulation (labial versus alveolar), following phoneme (vowel versus consonant), lexical status of the carrier stimulus (word versus nonword), and lexical competitor environment of the carrier stimulus. Results will be reported on the relative contribution of various durational, spectral, and energy measures to the voiced–voiceless distinction. The data showed that 25% of all tokens were produced without prevoicing. The prevoicing of the voiced plosive was omitted more often when the plosive was followed by a consonant than when it was followed by a vowel. Although both spectral and energy cues signal the voicing distinction, and although prevoicing is often omitted, VOT appears to be the primary cue to this distinction.

2aSC29. Using the Speech Transmission Index to predict the intelligibility of non-native speech. Sander J. van Wijngaarden, Herman J. M. Steeneken, Tammo Houtgast, and Adelbert W. Bronkhorst (TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands)

The calibration of the Speech Transmission Index (STI) is based on native speech, presented to native listeners. This means that the STI predicts speech intelligibility under the implicit assumption of fully native communication. In order to assess effects of both non-native production and non-native perception of speech, the intelligibility of short sentences was measured in various non-native scenarios, as a function of speech-to-noise ratio. Since each speech-to-noise ratio is associated with a unique STI value, this establishes the relation between sentence intelligibility and STI. The difference between native and non-native intelligibility as a function of STI was used to calculate a correction function for the STI for each separate non-native scenario. This correction function was applied to the STI ranges corresponding to certain intelligibility categories (bad–excellent). Depending on the proficiency of non-native talkers and listeners, the category boundaries were found to differ from the standard (native) boundaries by STI values up to 0.30 (on the standard 0–1 scale). The corrections needed for non-native listeners are greater than for non-native talkers with a similar level of proficiency. For some categories of non-native communicators, the qualification excellent requires an STI higher than 1.00, and therefore cannot be reached.

2aSC30. Acoustic correlates of Georgian ejectives. Tamra M. Wysocki (Dept. of Linguist., Univ. of Chicago, 1050 E. 59th St., Chicago, IL 60637, t-wysocki@uchicago.edu)

In this paper we present results from acoustic analysis of Georgian ejectives. Georgian, a language of the Kartvelian (South Caucasian) family, has a three-way opposition for voiced, voiceless aspirated and ejective stops. There have not been many acoustic studies of ejectives; those studies that discuss ejectives, such as by Lindau [J. Phonetics **12**, 147–155 (1984)], report cross-linguistic variation in timing between oral and glottal releases and the onset of a following vowel. In this paper we investigate acoustic correlates of Georgian ejectives and examine how these correlates are realized in two-member ejective clusters. Additionally, correlates of Georgian ejectives are compared to findings from previous studies and discussed in relation to cross-linguistic tendencies and variation. Data consist of field recordings of three native Georgian speakers producing single ejectives and ejective clusters in word-initial and word-medial (intervocalic) positions at a normal rate of speech. Acoustic analysis was completed using spectrograms and waveforms. Results indicate that, while there is variation within and among speakers, characteristics correlated with singleton Georgian ejectives include relative burst amplitude, noise quality following oral release, long, positive voice onset time, and a short period of creaky voice at the onset of a following vowel. Some of these characteristics differ from those of ejective clusters.

2aSC31. Japanese adult's and children's production and perception of English fricatives: A longitudinal study. Katsura Aoyama (Div. of Speech and Hearing Sci., Univ. of Alabama, Birmingham, CH20, 1530 3rd Ave. S., Birmingham, AL 35294, aoyama@uab.edu), Susan Guion (Univ. of Oregon, Eugene, OR 97403), Tsuneo Yamada (National Inst. of Multimedia Education, Chiba 261-0014, Japan), and Reiko Akahane-Yamada (ATR Human Information Processing Res. Labs., Kyoto 619-0288, Japan)

This study examined the production and perception of English fricatives by native Japanese (NJ) adults and children (16 per group, mean age=40 and 10 years), and age-matched native English (NE) adults and children (16 per group). The subjects were tested two times (T1, T2) 1 year apart. (At T1, the NJ subjects' mean length of residence in the U.S. was 0.5 year.) A picture-naming task was used to elicit the production of English words beginning with /s/ and /θ/, and intelligibility scores were obtained for both. The intelligibility scores of the NJ children but not adults improved significantly from T1 to T2. The NJ children obtained significantly lower scores than the NJ adults did at T1, but at T2 the adult–child difference was nonsignificant. The perception of /s/ and /θ/ was tested by a categorial discrimination task. Although the NJ adult's and children's scores improved from T1 to T2, the T1–T2 differences were nonsignificant. Thus, the results showed that the NJ children's production scores improved significantly from T1 to T2, while there was no significant change for the perception scores on the discrimination of /s/–/θ/. The relationship between production and perception in L2 speech learning will be discussed. [Work supported by NIH.]

2aSC32. Temporal features of Spanish-accented English. Amee Shah and Winifred Strange (Ph.D. Prog. in Speech and Hearing Sci., City Univ. of New York, 365 Fifth Ave., New York, NY 10016)

In this study we attempt to identify the acoustic parameters of Spanish accented English that give it the perception of being accented. Recordings of eight multisyllabic (3, 4, and 5 syllables) words spoken in sentences by 22 Spanish speakers of English and five native speakers of American English (AE) were analyzed for temporal acoustic differences. Segmental durations of Spanish productions, including differences of stressed–unstressed (s/u) vowel duration ratios, Voice Onset Time (VOT) and flap/

stop duration, differed from native productions. The sentences and the excised target words were given to native (AE) listeners (N=10) who judged the degree of accentedness on a 9-point scale. The degree of accentedness on target words correlated with perceived ratings on the eight sentences (i.e., global accentedness), rho=+0.82. Spearman rank-order correlations between s/u vowel duration ratios and native listener ratings of accentedness on target words varied from +0.04 to 0.68. The VOT duration of initial voiceless stops correlated positively with accentedness ratings for three of the four words (+0.26 to +0.35). Overall, results suggest that Spanish-accented English is characterized by significant temporal differences from native American English, and that these temporal differences contribute to the perception of accentedness as judged by native American English listeners.

2aSC33. Perception and analysis of Spanish accents in English speech. Cori Chism (Dept. of Foreign Lang., West Virginia Univ., Morgantown, WV 26506) and Norman Lass (West Virginia Univ., Morgantown, WV 26506)

The purpose of the present study was to determine what relates most closely to the degree of perceived foreign accent in the English speech of native Spanish speakers: intonation, vowel length, stress, voice onset time (VOT), or segmental accuracy. Nineteen native English speaking listeners rated speech samples from 7 native English speakers and 15 native Spanish speakers for comprehensibility and degree of foreign accent. The speech samples were analyzed spectrographically and perceptually to obtain numerical values for each variable. Correlation coefficients were computed to determine the relationship beween these values and the average foreign accent scores. Results showed that the average foreign accent scores were statistically significantly correlated with three variables: the length of stressed vowels ($r=-0.48$, $p=0.05$), voice onset time ($r=-0.62$, $p=0.01$), and segmental accuracy ($r=0.92$, $p=0.001$). Implications of these findings and suggestions for future research are discussed.

2aSC34. A cross-language study of vowel sounds produced with and without emphasis: Testing the theory of adaptive dispersion. Jessica F. Hay, Momoko Sato, Amy E. Coren, and Randy L. Diehl (Dept. of Psych., Univ. of Texas, Austin, TX 78712)

According to Lindbloms Theory of Adaptive Dispersion (TAD), the aim of talkers is to make phonological contrasts sufficiently distinctive to promote linguistic comprehension by the listener while minimizing the articulatory effort needed to achieve this degree of distinctiveness. When part of an utterance carries new—rather than given—information, it tends to be spoken with greater emphasis and clarity. In this study, several possible acoustic correlates of vowels in emphasized words were examined in American English, French, and Japanese in comparable phonetic and sentence contexts. These possible correlates include an expanded vowel space, greater vowel inherent spectral change, and a greater systematic variation in vowel length. Preliminary analyses suggest that the contrast-enhancing properties of emphasized vowels vary considerably across languages. [Work supported by NIDCD.]

TUESDAY MORNING, 4 JUNE 2002 BENEDUM ROOM, 9:25 TO 10:40 A.M.

Session 2aSPa

Signal Processing in Acoustics: Applications of Time Reversal Processing

Leon H. Sibul, Chair
Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

Contributed Papers

9:25

2aSPa1. Effectiveness of using time reverse signal processing for mitigating multipath in communications channels. Alan W. Meyer, James V. Candy, and Andrew J. Poggio (Lawrence Livermore Natl. Lab., 7000 East Ave., Livermore, CA 94550)

The use of time reverse signal processing (TRSP) to compensate for distortion in a communications channel is well established in recent literature. Our recent work quantifies the effectiveness of TRSP techniques *vis-à-vis* traditional adaptive equalization techniques. These quantitative measurements are developed using simulations and laboratory experiments. A rigorous mathematical development of the communications problem is first necessary to establish a consistent model of the communications channel, and various approaches for implementing TRSP and adaptive equalization as well. These techniques are then simulated and performance measures developed for a large set of conditions spanning an increasing number of multipaths and increasing SNR. Acoustic laboratory experiments are also conducted in a highly reverberant environment to confirm both our techniques and results. In this paper we report on these measures of effectiveness comparing TRSP vis-a-vis traditional adaptive equalization techniques for compensating for multipath in a communications channel. [Work performed under the auspices of the Department of Energy by the Lawrence Livermore National Laboratory under contract W-7405-Eng-48.]

9:40

2aSPa2. Broadband DORT applied to bistatic active sonar. David M. Fromm and Charles F. Gaumond (Naval Res. Lab., Washington, DC 20375-5350)

The decomposition of the time reversal operator (DORT) is a single frequency method of spatially isolating scatterers with a multiple-source/multiple-receiver system [C. Prada *et al.*, J. Acoust. Soc. Am. **99**, 2067–2076 (1996)]. The DORT technique applied to the active sonar problem has demonstrated the ability to isolate scatterers at various depths and ranges from distributed bottom reverberation for a set of monostatic or vertically bistatic source/receiver geometries. The sound channel propagation has several consequences for DORT analysis: (1) frequency fading prevents the source and receiving steering vectors for the first singular value from isolating the scatterer across the entire bandwidth of the signal; and (2) dispersion causes space–time spreading of the analysis window and complicates the back-propagation procedure. These effects become more pronounced as the technique is extended to fully bistatic source–receiver geometries and range-dependent environments. This paper presents the extension of the DORT technique to the full broadband, bistatic, active sonar problem, including refinements to the range-resolution/time-windowing and back-propagation techniques, and enhancement to detection and classification by exploiting the response for several singular values across the entire bandwidth of the signal. [Work supported by ONR.]

9:55

2aSPa3. Singular value spectra of transmission mode acoustic time reversal imaging. Sean K. Lehman (Lawrence Livermore Natl. Lab., L-154, 7000 East Ave., Livermore, CA 94550) and Anthony J. Devaney (CenSSIS (Ctr. for Subsurface Sensing and Imaging Systems), Northeastern Univ., 360 Huntington Ave., Boston, MA 02115)

We have formalized the theory of transmission mode time reversal imaging for arbitrary transducer geometry. Although it is not possible to implement such a system in practice, we have shown that transmission mode time reversal imaging is achievable *a posteriori* from the scattered field time series as measured at the receivers. We apply this theory to studying the singular value spectra of objects buried in the ground. We provide simulated finite-difference time-domain (FDTD) examples designed to model the case of geophysical vertical seismic profiling in which a series of sources are excited on the surface and the scattered fields are measured on a chain of receivers in a bore hole. A single, non-point scatterer can have multiple singular values dependent upon the scatterer's physical properties. We study the relationship between the singular value spectra and these physical properties.

10:10

2aSPa4. Time reversal in a layered medium. David H. Chambers and Alan W. Meyer (Lawrence Livermore Natl. Lab., P.O. Box 808 L-154, Livermore, CA 94551, chambers2@llnl.gov)

We investigate using a time reversal array to detect voids between two layers of materials. The array is positioned on the surface of a layered material so that acoustic pulses propagate in a direction roughly normal to the layers. These pulses are reflected from both the layer interface and small voids in the interface. The reflected pulses are time-reversed and sent back into the material. This process is iterated until it converges into an eigenstate of the time reversal operator associated with either a void or the interface. The eigenstates associated with the interface can be understood in terms of interactions between the array and its image in the interface. We look at the problem of detecting eigenstates associated with voids in the presence of the interface eigenstates. Strategies for classifying the eigenstates are derived by considering the simplified problem of a time reversal array on a single layer slab. [Work performed under the auspices of the Department of Energy by the Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.]

10:25

2aSPa5. Buried target detection using time reversal in inhomogeneous sediment. David Pierson and Thomas Drake (Marine, Earth and Atmospheric Sci., North Carolina State Univ., Raleigh, NC 27695)

Time reversal has been demonstrated to refocus acoustic waves in homogeneous media with multiple scatters. It has also been shown to work through a liquid–solid interface when both media are homogeneous. We will show that time reversal can also work to detect objects in inhomogeneous solids. By modeling a single transducer in water located 10 m above the water–solid interface we will show the location of the object can be isolated using an iterative time reversal method. Simulations are performed with the solid medium having a mean density of quartz sand varying spatially with horizontal correlation length 0.5 m and vertical correlation length 0.02 m. A sphere is buried in the sand having a density of iron in one test and a density closer to quartz sand in another. The sphere diameter was changed in each simulation ranging from 9 cm to 19 cm. The initial pulses were sinusoidal with frequencies of 10 kHz, 100 kHz, and 500 kHz. Results show improved localization of the sphere. It will also be shown that the iterative time reversal method goes to a response that is independent of initial pulse frequency as predicted by previous theories. [Research supported by ONR.]

TUESDAY MORNING, 4 JUNE 2002 BENEDUM ROOM, 10:50 TO 11:55 A.M.

Session 2aSPb

Signal Processing in Acoustics and Archives and History: History of Signal Processing in Acoustics

James V. Candy, Chair
Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, California 94551

Chair's Introduction—10:50

Invited Paper

2aSPb1. History of signal processing in acoustics. David I. Havelock (Natl. Res. Council, 1200 Montreal Rd., Ottawa, ON K1A 0R6, Canada, david.havelock@nrc.ca)

The Technical Committee on Signal Processing in Acoustics is the youngest of the ASA's 13 Technical Committees. It was formed as an Interdisciplinary Technical Group in Fall 1994 and was converted to a Technical Committee in Fall 2000. It has had over 120 technical sessions, cosponsoring sessions with all of the other Technical Committees. Signal processing first appeared as a distinct PACS classification in 1964, but significant papers on signal processing appear in JASA well before this. The popularization of the fast Fourier transform and developments in digital computer technology spurred research and education in signal processing. In more recent years, advances in digital technology have allowed digital signal processing to advance beyond ''off-line'' simulation and analysis to embedded components of everyday devices and systems. As signal-processing capabilities advance, increased opportunities for cross-disciplinary applications can be expected. Accordingly, an important role for the Technical Committee on Signal Processing in Acoustics will be to enhance this interchange within the ASA. The history of signal processing in acoustics will be presented through examples from various areas of acoustics.

Session 2pAA

Architectural Acoustics and Noise: Computer Modeling in Acoustical Consulting I

Lily M. Wang, Chair
Architectural Engineering, University of Nebraska, 200B Peter Kiewit Institute, 1110 South 67th Street, Omaha, Nebraska 68182-0681

Chair's Introduction—1:00

Invited Papers

1:05

2pAA1. Computer modeling in the practice of acoustical consulting: An evolving variety of uses from marketing and diagnosis through design to eventually research. Gary S. Madaras (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774)

The use of computer modeling as a marketing, diagnosis, design, and research tool in the practice of acoustical consulting is discussed. From the time it is obtained, the software can be used as an effective marketing tool. It is not until the software basics are learned and some amount of testing and verification occurs that the software can be used as a tool for diagnosing the acoustics of existing rooms. A greater understanding of the output types and formats as well as experience in interpreting the results is required before the software can be used as an efficient design tool. Lastly, it is only after repetitive use as a design tool that the software can be used as a cost-effective means of conducting research in practice. The discussion is supplemented with specific examples of actual projects provided by various consultants within multiple firms. Focus is placed on the use of CATT-Acoustic software and predicting the room acoustics of large performing arts halls as well as other public assembly spaces.

1:35

2pAA2. Recent emphasis of computer modeling in architectural acoustics consulting. Ted Pyper, Derrick Knight, and Andrew Schmidt (Jaffe Holden Acoust., 114A Washington St., Norwalk, CT 06854, tpyper@jhacoustics.com)

Jaffe Holden Acoustics has been designing successful performing arts spaces for over four decades. Within recent years, the use of computer modeling programs, such as CATT Acoustic, has gained wide acceptance as a design tool in acoustical consulting. JHA has found computer modeling to be useful in refining some specific design issues. Models will be presented for Zankel Recital Hall at Carnegie Hall, Kennedy Center Concert Hall, Marion Oliver McCaw Hall in Seattle, and the University of Connecticut band rehearsal room. The purpose and application of the models will be discussed, along with the conclusions drawn from each study.

2:05

2pAA3. Computer modeling/auralization: Can these computer aids assist architects and owners in making proper project decisions? Robert C. Coffeen (School of Architecture and Urban Design, Marvin Hall, The Univ. of Kansas, Lawrence, KS, coffeen@ku.edu)

Computer modeling of architectural spaces and subsequent auralization provides the acoustical consultant with information which is not easily obtained by other techniques and methods, and which leads to better acoustical designs. But, what information does modeling/auralization produce which will help the architect and facility owner make proper decisions relating to room acoustics and electro-acoustics when such decisions have economic and architectural implications? This paper will present several examples of how modeling/auralization will assist in this quest for acoustical understanding and suitable project decisions.

2:35

2pAA4. Applications of computer modeling at Wrightson, Johnson, Haddon & Williams, Inc. James A. Johnson and Benjamin C. Seep (WJHW, Inc., 4801 Spring Valley Rd., #300, Dallas, TX 75244, jjohnson@wjhw.com)

Computer modeling has become useful as an investigative tool and as a client communication and explanation tool in the field of acoustical consulting. A variety of in-house developed and commercially available applications is in constant use at the firm of Wrightson, Johnson, Haddon & Williams. Examples likely to be demonstrated (depending on time) include use of digital filtering for building exterior noise reduction comparisons, a shell isolation rating (SIR) model, simple sound barrier programs, an HVAC spreadsheet, a visual sightline modeling tool, specular sound reflections in a semicircular arc, and some uses of CATT-acoustic auralizations.

3:05–3:20 Break

3:20

2pAA5. Utilizing computer models for optimizing classroom acoustics. Jennifer M. Hinckley and Carl J. Rosenberg (Acentech, Inc., 33 Moulton St., Cambridge, MA 02138, jhinckley@acentech.com)

The acoustical conditions in a classroom play an integral role in establishing an ideal learning environment. Speech intelligibility is dependent on many factors, including speech loudness, room finishes, and background noise levels. The goal of this investigation was to use computer modeling techniques to study the effect of acoustical conditions on speech intelligibility in a classroom. This study focused on a simulated classroom which was generated using the CATT-acoustic computer modeling program. The computer was utilized as an analytical tool in an effort to optimize speech intelligibility in a typical classroom environment. The factors that were focused on were reverberation time, location of absorptive materials, and background noise levels. Speech intelligibility was measured with the Rapid Speech Transmission Index (RASTI) method.

3:50

2pAA6. A case-study comparison of computer modeling and scale modeling in acoustics consulting. Paul T. Calamia (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

As an alternate or compliment to computer models, acoustics consultants often make use of scale models to evaluate the efficacy of architectural designs. The intention of this paper is to compare the two modeling approaches, using one or more case studies, to explore the pros and cons of each. Topics of comparison will include cost, geometric representations, effective bandwidths, propagation phenomena (e.g., diffraction), simulation of material properties, and auralization. Where possible, measured data from existing spaces will be presented to provide a reference for the modeled data.

4:20

2pAA7. Practical considerations in using audible simulation technology. Thomas Tyson and Kenneth Jacob (Bose Professional Systems Div., The Mountain, Framingham, MA 01701)

Ideally, consultants and clients alike would like to hear sound from computer models that is indistinguishable from the sound of the finished project. No one, however, would claim perfection in computer modeling, and differences between the model and the finished project certainly remain. How can practitioners in the field know exactly what these differences are, and more importantly, how they may affect the outcome of their work? Said a different way, how much can they and their clients trust what they hear from their computer models? In the work presented here, an approach used to answer these questions in the case of Boses Modeler and Auditioner programs is given. The approach is explained using: (1) laboratory results that focus on the critical dimension of speech intelligibility and (2) comparisons of actual system installations from the past seven years that were modeled and presented to end users using the Auditioner system.

TUESDAY AFTERNOON, 4 JUNE 2002 LE BATEAU ROOM, 1:00 TO 2:35 P.M.

Session 2pAB

Animal Bioacoustics: Session on Marine Mammal Bioacoustics in Honor of William E. Evans II

Ann E. Bowles, Cochair
Hubbs-Sea World Research Institute, 2595 Ingraham Street, San Diego, California 92109

Jeanette A. Thomas, Cochair
Laboratory of Sensory Biology, Western Illinois University, 3561 60th Street, Moline, Illinois 61265

Chair's Introduction—1:00

Contributed Papers

1:05

2pAB1. A window on perception: Response times of odontocete cetaceans in audiometric tests. Diane J. Blackwood (Marine Acoust. Lab., Texas A&M Univ. and 3027 Macaulay St., San Diego, CA 92106, blackwood@cls.org), Sam H. Ridgway (Navy Marine Mammal Prog., San Diego, CA 92152), and William E. Evans (Marine Acoust. Lab., Texas A&M Univ.)

A standard psychometric measurement is response time, the interval elapsing between a stimulus and a response. While studies of response time have been published for humans and other terrestrial mammals, this study marks the first report of response times for odontocete cetaceans at threshold in an audiometric task. Two white whales (*Delphinapterus leucas*) and four Atlantic bottlenose dolphins (*Tursiops truncatus*) were given audiometric tests to determine masked hearing thresholds. Animals were tested at 26 frequencies over a range from 200 Hz to 100 kHz using pure tones. The test tone amplitudes covered a range of 20 dB *re* 1 microPascal including the hearing threshold of the animal at that frequency. Hearing thresholds varied from 87.5 dB to 125.5 dB depending on frequency, masking noise intensity and individual animal. Data was analyzed to determine characteristic relationships between response time and amplitude of test tone for each frequency and animal. The two whales responded

significantly slower (640 ms, 0.001) than the four dolphins (430 ms). As in terrestrial animals, reaction time became shorter as stimulus strength increased. At threshold, median response time across frequencies within each animal varied about 150 ms.

1:20

2pAB2. What can be learned from one of nature's most advanced biosonars: Discussion on Bottlenose dolphins echolocation waveforms with respect to echolocation tasks in shallow water. Marc P. Olivieri (Broadband Sensor Div., GORCA Technologies, Inc., 300 W. Rte. 38, Moorestown, NJ 08057, olivieri@gorca.com)

This paper presents recent findings on the biosonar echolocation waveforms of Bottlenose dolphins (*Tursiops truncatus*). The complex spectral shape of the waveforms (bimodal spectrum) and the dynamics of the click trains are discussed. It is shown that these waveforms have not only the smallest possible time-bandwidth product, but that they are in fact minimum phase waveforms (i.e., the phase is a function of the autospectrum). The characteristics of the waveforms including time/frequency distribution, autocorrelation, and the bimodal spectral structure of these waveforms are discussed. An interpretation of these features is proposed with respect to echolocation tasks in shallow water. It is shown how these waveforms may be used for optimizing both near range resolution and sidelobe levels in the processed echo waveforms. Examples are shown on how GORCA Technologies, Inc. (GTI) exploits such features in the development of a biomimetic SONAR. Finally, this paper presents examples of GTIs proprietary broadband adaptive algorithm performance, using animal echolocation data. In the data used for this study, the animal performed a three-alternative match-to-sample task (data courtesy of the biosonar program at SPAWAR San Diego). [Work sponsored by ONR.]

1:35

2pAB3. Nocturnal location and activity of Pacific coast bottlenose dolphins determined by a sonobuoy array. Eric S. Howarth and R. H. Defran (Cetacean Behavior Lab., San Diego State Univ., 5500 Campanile Dr., San Diego, CA 92182, ehowarth@mail.sdsu.edu)

Pacific coast bottlenose dolphins (*Tursiops truncatus*) move offshore, presumably to feed as the sun sets [Hanson and Defran (1993), Ward (1998), Day (1998)] and are most often found inshore on the surfline at sunrise. The purpose of this research was to document their activity and location during nocturnal hours. A sonobuoy array was deployed in a triangle configuration centered approximately 500 m from the mean low low water line at Torrey Pines State Beach, CA. Surveys were conducted for 11 nights between October 1999 and November 1999 and for 21 nights between August 2000 and December 2000. Dolphin vocalizations were detected on 26 of 33 nights.

1:50

2pAB4. Specialization of the auditory processing in harbor porpoise, characterized by brain-stem potentials. Nikolay G. Bibikov (N. N. Andreyev Acoust. Inst., Shvernik St., 4, Moscow 117036, Russia, nbibikov@hotmail.com)

Brain-stem auditory evoked potentials (BAEPs) were recorded from the head surface of the three awaked harbor porpoises (Phocoena phocoena). Silver disk placed on the skin surface above the vertex bone was used as an active electrode. The experiments were performed at the Karadag biological station (the Crimea peninsula). Clicks and tone bursts were used as stimuli. The temporal and frequency selectivity of the auditory system was estimated using the methods of simultaneous and forward masking. An evident minimum of the BAEPs thresholds was observed in the range of 125–135 kHz, where the main spectral component of species-specific echolocation signal is located. In this frequency range the tonal forward masking demonstrated a strong frequency selectivity. Off-response to such tone bursts was a typical observation. An evident BAEP could be recorded up to the frequencies 190–200 kHz, however, outside the acoustical fovea the frequency selectivity was rather poor. Temporal resolution was estimated by measuring BAER recovery functions for double clicks, double tone bursts, and double noise bursts. The half-time of BAERs recovery was in the range of 0.1–0.2 ms. The data indicate that the porpoise auditory system is strongly adapted to detect ultrasonic closely spaced sounds like species-specific locating signals and echoes.

2:05

2pAB5. Whale contribution to long time series of low-frequency oceanic ambient sound. Rex K. Andrew, Bruce M. Howe, and James A. Mercer (Appl. Phys. Lab., 1013 NE 40th St., Seattle, WA 98040, randrew@apl.washington.edu)

It has long been known that baleen (mainly blue and fin) whale vocalizations are a component of oceanic ambient sound. Urick reports that the famous ''20-cycle pulses'' were observed even from the first Navy hydrophone installations in the early 1950's. As part of the Acoustic Thermometry Ocean Climate (ATOC) and the North Pacific Acoustic Laboratory (NPAL) programs, more than 6 years of nearly continuous ambient sound data have been collected from Sound Surveillance System (SOSUS) sites in the northeast Pacific. These records now show that the average level of the ambient sound has risen by as much as 10 dB since the 1960's. Although much of this increase is probably attributable to manmade sources, the whale call component is still prominent. The data also show that the whale signal is clearly seasonal: in coherent averages of year-long records, the whale call signal is the only feature that stands out, making strong and repeatable patterns as the whale population migrates past the hydrophone systems. This prominent and sometimes dominant component of ambient sound has perhaps not been fully appreciated in current ambient noise models. [Work supported by ONR.]

2:20

2pAB6. Near-bottom hydrophone measurements of ambient noise and sperm whale vocalizations in the northern Gulf of Mexico. Joal Newcomb, Robert Fisher, Robert Field, Altan Turgut (Naval Res. Lab.-Stennis, Stennis Space Center, MS 39576 newcomb@nrlssc.navy.mil), George Ioup, Juliette Ioup (Univ. of New Orleans, New Orleans, LA 70148), Grayson Rayborn, Stan Kuczaj, Jerald Caruthers, Ralph Goodman (Univ. of Southern MS, Hattiesburg, MS 39402), and Natalia Sidorovskaia (Univ. of Louisiana at Lafayette, Lafayette, LA 70504)

Three bottom-moored hydrophones, 50 m above the bottom, were placed on a downslope line, ending at the largest concentration of sperm whale sightings in the northern Gulf of Mexico, in 600 m, 800 m, and 1000 m water depths. These depths were chosen after upslope propagation modeling, using historical databases, showed transmission losses greater than 110 dB at hydrophones near the bottom in water shallower than 600 m for a 500 m deep source at the 1000 m contour. These autonomously recording hydrophones were environmental acoustic recording system (EARS) buoys obtained from the Naval Oceanographic Office. They were capable of recording signals up to 5500 Hz continuously for 36 days and were deployed from July 17 through August 21. During this period a major marine mammal exercise was being conducted at the surface by the Minerals Management Service and the National Marine Fisheries Service, with other government and university scientists, in which temporary acoustic recording devices were attached to the whales and the whales were monitored by a surface towed array. Our near-bottom measurements of ambient noise and sperm whale vocalizations are discussed and compared to those surface and on-whale measurements. [Research supported by ONR.]

Session 2pAO

Acoustical Oceanography: Acoustic Remote Sensing of the Ocean Environment

David R. Palmer, Chair

Atlantic Oceanographic and Meteorological Laboratory, National Oceanographic and Atmospheric Administration, 4301 Rickenbacker Causeway, Miami, Florida 33149

Contributed Papers

3:15

2pAO1. Estimating Europa's internal structure with ambient noise. Sunwoong Lee, Michele Zanolin, Aaron M. Thode, and Nicholas C. Makris (MIT, 77 Massachusetts Ave., Cambridge, MA 02139, makris@mit.edu)

Jupiter's moon, Europa, is believed to possess a vast liquid ocean beneath a thin ice surface, although the thickness of the ice shell and depth of the ocean still remain poorly constrained. The authors have previously discussed the possibility of using ambient noise to probe Europa's interior [Makris *et al.*, Eos. Trans. Am. Geophys. Union **82**(47), P22B-0552 (2001)]. Here the analysis is refined and extended. The probability density for high source level or ''big bang'' events from tidally induced surface cracks and meteor impacts is estimated. Such events are important because they are most likely to have returns from internal Europan strata that stand above the ambient noise level. Travel-time curves for various thicknesses are also presented and used to demonstrate that echo-sounding and tomographic techniques can be applied to estimate thickness of the ice layer, depth of the ocean, and internal temperature structure. Convective ice can also be detected by spectral analysis of surface waves, since significant contrast in shear wave speed is expected in the partial-melt-ice regime. Spectral analysis is also applied to analyze flexural modes of a thin ice shell, to estimate ice-shell thickness, and to potentially study the sediment layer on the seafloor.

3:30

2pAO2. Full field spatial correlation of range dependent surface generated noise in a stratified ocean with application to hurricane sensing. Joshua D. Wilson and Nicholas C. Makris (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

Much of the ambient noise in the ocean is generated by wind at the ocean surface. In the case of a hurricane, the wind speed, and therefore the source level, is not constant over the ocean surface. The wind speed in a hurricane can range from 0 knots at the eye to 140 knots at the eye wall where the eye and the eye wall are typically separated by only 10 kilometers. A spectral model for the spatial correlation of range dependent surface generated noise based on wave number integration is developed. This model accurately represents the acoustic field directly beneath the hurricane and at arbitrarily long ranges. It is used to illustrate the dramatic variation in the noise field caused by range dependent features like the eye and eye wall. Results of this model are shown for hurricanes in shallow and deep water environments, including Bay of Bengal and Mid-Atlantic regions. This model is also compared with a previous normal mode model [Perkins *et al.*, J. Acoust. Soc. Am. **93**, 739–752 (1993)]. This comparison shows that at long range the two models agree but at close range the wave number integration model is necessary to accurately represent the field.

3:45

2pAO3. Accuracy limits on rapid assessment of gently varying bathymetry. B. Edward McDonald (U.S. Naval Res. Lab., Washington, DC 20375, mcdonald@sonar.nrl.navy.mil) and Charles Holland (Penn State Univ., State College, PA 16804)

Accuracy limits for rapidly probing shallow water bathymetry are investigated as a function of bottom slope and other relevant parameters. The probe scheme [B. E. McDonald and Charles Holland, J. Acoust. Soc. Am. **110**, 2767 (2001)] uses a time reversed mirror (TRM) to ensonify a thin annulus on the ocean bottom at ranges of a few km from a vertical send/ receive array. The annulus is shifted in range by variable bathymetry (perturbation theory shows that the focal annulus experiences a radial shift proportional to the integrated bathymetry along a given azimuth). The range shift implies an azimuth-dependent time of maximum reverberation. Thus the reverberant return contains information that might be inverted to give bathymetric parameters. The parameter range over which the perturbation result is accurate is explored using the RAM code for propagation in arbitrarily range-dependent environments. [Work supported by NRL.]

4:00

2pAO4. Revisiting Buldyrev's theory of the ''axial wave'' after 30 years. David R. Palmer (NOAA/Atlantic Oceanogr. and Meteorological Lab., 4301 Rickenbacker Cswy., Miami, FL 33149, david.r.palmer@noaa.gov), Natalie S. Grigorieva, and Gregory M. Fridman (St. Petersburg State Marine Tech. Univ., St. Petersburg, Russia)

We review and summarize some of the more interesting aspects of Buldyrev's theory of the axial wave as it relates to long-range propagation. Just as with the experimental observations, the theory describes the near-axial acoustic signal in terms of early, geometrical-like arrivals followed by an incoherent arrival that cannot be described using geometrical acoustics. This incoherent arrival, or axial wave, results from a large number of waves which interfere with one another because of the presence of caustic structures along the waveguide axis. It is this interference which prevents the waves from being described in terms of geometrical acoustics. As the propagation range increases, the nature of the axial wave changes because the number of caustics increases and the region of interference associated with each one also increases. For a normal waveguide (axial energy arrives last) the axial wave ''eats'' a geometrical arrival as each new caustic is encountered along the propagation path. A range is ultimately reached beyond which no geometric arrivals are present in the signal. In underwater acoustics this maximum range is probably unphysically large. For an abnormal waveguide (axial energy arrives first), the axial wave ''spits'' out geometrical arrivals as the range increases.

4:15

2pAO5. Investigation of near-axial interference effects in long-range acoustic propagation in the ocean. Natalie S. Grigorieva and Gregory M. Fridman (Dept. of Appl. Mathematics and Mathematics Modeling, St. Petersburg State Marine Tech. Univ., 3 Lotsmanskaya Str., St. Petersburg 190008, Russia, nsgrig@natalie.spb.su)

The observed time-of-arrival patterns from a number of long-range ocean acoustic propagation experiments show early geometrical-like arrivals followed by a crescendo of energy that propagates along the sound-channel axis and is not resolved into individual arrivals. The two-

dimensional reference point source problem for the parabolic index of refraction squared is investigated to describe in a simple model case the interference of near-axial waves which resulted in forming the so-called axial wave and propose a formula for the axial wave in more general cases. Using the method proposed by Buldyrev [V. Buldyrev, Tr. Mat. Inst. Steklov **115**, 78–102 (1971)], the integral representation for the exact solution is transformed in such a way to extract ray summands corresponding to rays radiated from the source at angles less than a certain angle, the axial wave, and a term corresponding to the sum of all the rays having launch angles greater than the indicated angle. Numerical results for the axial wave and the last term are obtained for parameters corresponding to long-range ocean acoustic propagation experiments. The generalization of the obtained formula for the axial wave to the case of an arbitrary range-independent sound speed is given and discussed. [Work supported by VSP Grant No. N00014-01-4003.]

TUESDAY AFTERNOON, 4 JUNE 2002 BRIGADE ROOM, 1:30 TO 3:45 P.M.

Session 2pEA

Engineering Acoustics: Noise and Other Engineering Acoustics Topics

Ahmet Selamet, Chair
Center for Automotive Research, The Ohio State University, 930 Kinnear Road, Columbus, Ohio 43212

Chair's Introduction—1:30

Contributed Papers

1:35

2pEA1. The ambiguity of the acoustic source distributions for divergent source fields, with application to aerodynamic noise. Alan Powell (Dept. of Mech. Eng., Univ. of Houston, TX 77204-4006)

A *given* uniform applied line force of linear strength F_1 falls very rapidly to zero about the ends of its length L_1. The wave equation source term is *monopolelike*, $\partial F_1/\partial y_1$. The radiation is *only* from the opposite energy producing monopoles $\int(\partial F_1/\partial y_1)dy_1 = \pm F_1$ at its extremities, forming a dipole L_1F_1. The divergence theorem transforms this into the *exactly* equivalent *dipole distribution* of strength F_1, total strength F_1L_1, now the "obvious" energy source for the *given* excitation. Analogously, a volume $L_1L_2L_3$ of a *quadrupole* distribution of uniform strength, say $\rho u_1 u_2$ for an aerodynamic lateral quadrupole, also radiates like opposing line *dipoles* $\int(\partial \rho u_1 u_2/\partial y_2)dy_2 L_3 = \pm \rho u_1 u_2 L_3$ along the L_1 sides L_2 apart (for lateral quadrupole $L_2 \rho u_1 u_2 L_1 L_3$), replaceable by *monopoles* $\int(\partial^2 \rho u_1 u_2/dy_1\, \partial y_2)dy_1\, dy_2 L_3$ at the corners of the face L_1L_2 (where $\partial^2 \rho u_1 u_2/\partial y_1\, \partial y_2 \neq 0$), yielding lateral quadrupole $L_1L_2\rho u_1 u_2 L_3$. With *exactly* the same fields, which can be considered physically realistic, like the force? The situation appears quite ambiguous; although note the fictitiousness of the monopoles, the appeal of $\partial \rho u_1 u_2/\partial y_2$ appearing in momentum equation as an applied force, while the energy loss by all elements, *directly*, of the sound producing flow may perhaps be best associated with the quadrupole distribution.

1:50

2pEA2. The sound production characteristics of high-frequency impinging jet tones. Brenda S. Henderson (Mech. Eng. Dept., Kettering Univ., Flint, MI 48504-4898), James Bridges, and Mark P. Wernet (NASA Glenn Res. Ctr., Cleveland, OH 44135)

An experimental investigation into the production of high-frequency tones by a supersonic impinging jet using phase-locked shadowgraph photography and high-resolution digital particle image velocimetry (DPIV) is presented. High-frequency tones are often produced for nozzle-to-plate spacings less than two nozzle exit diameters and for nozzle pressure ratios (NPR) less than approximately 3.4, where NPR is equal to the stagnation pressure divided by the pressure at the nozzle lip. The high-frequency tones have lower amplitudes and less impulsive wavefronts than the more dominant low-frequency impinging tones and are presumably produced by a significantly different sound production mechanism. Periodic motion of the jet appears to be confined to the peripheral regions of the jet with no significant motion of the flow in the central regions of the jet behind the standoff shock wave. Unlike the jet motion associated with low-frequency impinging jet tones, recirculating regions are not formed and swept downstream behind the standoff shock wave and stationary stagnation regions are also not observed.

2:05

2pEA3. Shallow cavity flow tone experiments: Onset of locked-on states. D. Rockwell, J.-C. Lin, P. Oshkai, M. Reiss (Lehigh Univ., 19 Memorial Dr. W., Bethlehem, PA 18015-3085), and M. L. Pollack (Lockheed Martin Corp., Schenectady, NY 12301)

Fully turbulent inflow past a shallow cavity is investigated for the configuration of an axisymmetric cavity mounted in a pipe. Emphasis is on conditions giving rise to coherent oscillations, which can lead to locked-on states of flow tones in the pipe–cavity system. Evaluation of pressure gradients on plan views of three-dimensional representations allows extraction of the frequencies of the instability (Strouhal) modes of the cavity oscillation. These frequency components are correlated with traditional models originally formulated for cavities in a free stream. In addition, they are normalized using two length scales; inflow boundary-layer thickness and pipe diameter. These scales are consistent with those employed for the hydrodynamic instability of the separated shear layer, and are linked to the large-scale mode of the shear layer oscillation, which occurs at relatively long cavity length. The foregoing considerations provide evidence that pronounced flow tones can be generated from a fully turbulent inflow at very low Mach number. These tones can arise even for the extreme case of a cavity having a length over an order of magnitude longer than its depth. Suppression of tones is generally achieved if the cavity is sufficiently shallow.

2:20

2pEA4. Wind tunnel measurements of windscreen performance. Edward R. Maniet, Jr. (Textron Systems, 201 Lowell St., Rm. 3106, Wilmington, MA 01887, emaniet@systems.textron.com)

Microphone windscreens are routinely used in outdoor acoustic measurements to reduce wind noise pickup. Characterization of windscreen performance outdoors with real wind has several drawbacks notably that the test conditions cannot be controlled. Test methodology has been de-

veloped that provides controlled, repeatable measurements of microphone windscreen performance in a laboratory setting. Wind noise measurements are performed using a high-speed/laminar-flow, low-noise wind tunnel that incorporates a large anechoic chamber. The wind tunnel is modified to produce a turbulent flow into which the microphone and windscreen under evaluation are placed. The turbulent velocity spectrum is measured using multi-axis hot-wire anemometers and compared to outdoor data to locate the best position in the turbulent flow to place the test article. Comparative performance measurements of several windscreen designs are presented.

2:35–2:45 Break

2:45

2pEA5. Acoustic characteristics of engine air cleaner boxes. Iljae Lee, Ahmet Selamet (Ctr. for Automotive Res., The Ohio State Univ., 930 Kinnear Rd., Columbus, OH 43212, selamet.1@osu.edu), and Sean Li (DaimlerChrysler Corp.)

Air cleaner boxes placed on the induction system of engines play a critical role in the attenuation of airborn sound. The acoustic performance of such configurations is investigated experimentally and computationally in the absence of mean flow. The transmission loss of a fabricated rectangular box (prototype) with various inlet and outlet locations and extensions is predicted by a three-dimensional boundary element method. The predictions are then compared with experimental data obtained from an impedance tube setup. The results demonstrate that: (1) the variations of locations and extension lengths of inlet and outlet affect the acoustic behavior of the prototype at higher frequencies, as expected; (2) the transmission loss characteristics at low frequencies may be estimated by a one-dimensional analysis; and (3) the boundary element method is effective throughout the entire frequency range of interest in predicting the acoustic performance of air cleaner boxes with asymmetric shape and/or inlet and outlet extensions.

3:00

2pEA6. A parametric study of location effects on the sidebands of Helmholtz resonators using boundary element analysis. Christopher E. Shaw, David J. Moenssen, and John D. Kostun (Visteon Corp., 15200 Commerce Dr., North Dearborn, MI 48120, cshaw2@visteon.com)

An important noise source used to tune automotive interior sound levels and quality is the sound which radiates from the open end of the engine induction system—induction noise. Understanding this source and its tuning can be vital to produce the customer-desired sound inside the vehicle. The most traditional method to attenuate (or tune) induction noise is through the addition of passive devices, the most common of which is the Helmholtz resonator. An unfortunate side effect to such a device is that narrow bands of frequencies higher and lower than the bandwidth of the resonator, called sidebands, tend to be amplified by the device. The subject of this study is to qualify the effect of changing resonator location on these sidebands. This is studied by determining the system frequency response to a white noise input by performing boundary element analysis on a simplified air induction system. The location of the device is varied, and the magnitude and frequency shift of the sidebands are studied. Experimental validation is also provided for select models.

3:15

2pEA7. An experimental/numerical study of nonlinear standing waves in resonantors. Joshua R. Finkbeiner, Xiaofan Li, Ganesh Raman (Illinois Inst. of Technol., 10 W. 32nd St., E1, Chicago, IL 60616), Christopher Daniels, and Bruce Steinetz (NASA Glenn Res. Ctr., 21000 Brookpark Rd., M.S. 23-3, Brook Park, OH 44135)

The generation of shock-free high-amplitude pressure waves in resonators promises to become a powerful and useful technology. Lawrenson *et al.* [J. Acoust. Soc. Am. Aug. (1998)] described the generation of shock-free high-amplitude pressure using large equipment and resonators to produce the reported effects. An attempt is made to generate shock-free high-amplitude pressure waves using relatively small resonators excited by a small linear motor. Ambient air is used as the working fluid. A small cylindrical resonator is tested resulting in the lack of the expected shocked waveform while a larger model of the same shape exhibits shocking behavior. A small conical resonator produces shock-free pressure waves at resonance, but the amplitude of these waves is small. A larger cone resonator model produces shock-free pressure waves of higher amplitude. A large horn-cone resonator also produces shock-free high-amplitude pressure waves. A numerical model solves the governing equations using a shooting method. The numerical results are compared to the experimental results. The difficulties caused by structural resonances, which are separate from the desired fluid resonances and caused by approaching the natural frequency of the linear motor/resonator structure, are shown and discussed.

3:30

2pEA8. Resonant acoustic measurement of vapor phase transport phenomenon in porous media. Richard Schuhmann (Dept. of Civil & Environ. Eng., Penn State) and Steven Garrett (Grad. Prog. in Acoust., Penn State)

Diffusion of gases through porous media is commonly described using Fick's law and is characterized by a gas diffusion coefficient modified by a media-specific tortuosity parameter. A phase-locked-loop resonance frequency tracker [J. Acoust. Soc. Am. **108**, 2520 (2000)] has been upgraded with an insulated copper resonator and a bellows-sealed piston instrumented with an accelerometer. Average system stability (temperature divided by frequency squared) is about 180 ppm. Glass-bead-filled cores of different lengths are fitted into an o-ring sealed opening at the top of the resonator. The rate at which the tracer gas is replaced by air within the resonator is controlled by the core's diffusion constant. Mean molecular weight of the gas mixture in the resonator is determined in real time from the ratio of the absolute temperature to the square of the fundamental acoustic resonance frequency. Molecular weight of the gas mixture is determined approximately six times per minute. Changes in the gas mixture concentration are exponential in time (within 0.1%) over nearly two decades in concentration. We will report diffusion constants for two different sizes of glass beads, in samples of five different lengths, using two different tracer gases, to establish the validity of this approach. [Work supported by ONR.]

Session 2pMUa

Musical Acoustics: Free Reed Instruments: Historical Perspective and Recent Research

James P. Cottingham, Chair
Physics Department, Coe College, Cedar Rapids, Iowa 52402

Chair's Introduction—1:00

Invited Papers

1:05

2pMUa1. From Kratzenstein to Wheatstone: Episodes in the early history of free reed acoustics. James P. Cottingham (Phys. Dept., Coe College, Cedar Rapids, IA 52402, jcotting@coe.edu)

In 1780 C. G. Kratzenstein published a paper in St. Petersburg describing a machine which produced vowel sounds using free reeds with resonators of various shapes. This marks a convenient, if arbitrary, starting point for the history of the free reed musical instruments of European origin. These instruments developed rapidly, and by 1850 the accordion, concertina, harmonica, reed organ, and harmonium all had been invented and developed into more or less final form. A key figure in this period is Charles Wheatstone, who not only published papers on acoustical research but was also an inventor and commercially successful manufacturer of musical instruments, most notably the Wheatstone English concertina. Much of Wheatstone's research in acoustics and almost all of his work as an inventor of musical instruments involved free reeds. This paper presents some episodes in the development of the free reed instruments and some examples of acoustical research involving free reeds during the 18th and 19th centuries.

1:35

2pMUa2. On Weber's theory of a free reed coupled to a resonator. Jonas Braasch (Institut für Kommunikationsakustik, Ruhr-Universität Bochum, 44780 Bochum, Germany)

In the 1820s Wilhelm Weber investigated the effect of resonator length on the frequency of a free reed coupled to a resonator. His aim was to build a device to generate a reference frequency since he was not satisfied with the accuracy of tuning forks. Weber's goal was to compensate the small dependence of frequency on blowing pressure and temperature by coupling a resonator to the free reed. Besides conducting a series of experiments Weber was the first to formulate a theory to estimate the frequency of a free-reed/resonator system. His theory is based on Euler's solution for two coupled pendulums with the free reed and the air column of the resonator as the two oscillators. Although some of his assumptions were not quite correct, Weber was able to predict some of the experimental results quite accurately by assuming that the oscillators are synchronized. However, because the free reed by itself was not frequency dependent on blowing pressure and temperature in his theory, he was not able to find a solution for compensating the frequency of the free reed using a resonator. In the presentation Weber's work will be introduced and compared to present theories.

2:05

2pMUa3. Measurements of reed vibration and pressure variation of the sho, the Japanese mouth organ. Takafumi Hikichi and Naotoshi Osaka (NTT Commun. Sci. Labs., NTT Corp., 3-1, Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan, hikichi@brl.ntt.co.jp)

Measurements of reed vibration and pressure vibration of the sho were carried out. Reed displacement, sound pressure at both sides of the reed, and radiated sound pressure at the open end of the pipe were measured using an experimental sho model made from an acrylic pipe and a metal reed. The kashira, or cavity, was also made from an acrylic box, where the pipe is mounted so that the reed vibration can be measured by a laser displacement sensor. The measurement results show that reeds vibrate as a sinusoidal, and that, in contrast with earlier results obtained by experiments on harmonium reeds [J. P. Cottingham, C. J. Lilly, and C. H. Reed, 137th meeting of the ASA and the 2nd Convention of the EAA, pp. 14–19 (1999)], the amplitude of reed vibration increases with increasing blowing/drawing pressure. The sound pressure inside the tube shows peaks when the reed reaches its maximum displacement, and the sound pressure oscillates twice when the reed oscillates once. Further, sound pressure inside the kashira, i.e., at the upstream of the reed, shows high frequency oscillation compared with reed vibration. This result implies that the shape and volume of the kashira may significantly affect the sound characteristics.

Contributed Paper

2:35

2pMUa4. Does the European free reed really originate from Asia? Jonas Braasch (Institut für Kommunikationsakustik, Ruhr-Universität Bochum, 44780 Bochum, Germany) and Christian Ahrens (Ruhr-Universität Bochum, 44780 Bochum, Germany)

Free reeds can be found in European instruments like the accordion and reed organ. Christian Gottlob [Theophil] Kratzenstein was the first scientist in Europe to use a free reed when he built his speech-imitating machine, and it is believed throughout literature that he copied the mechanism of a free reed from the Asian mouth organ. However, after analyzing Kratzenstein's original work (*Tentamen Resolvendi Problema ab Academia Scientiarum Imperiali Petropolitana ad Annum 1780 Publicae Propositum*) it was rather concluded that his free-reed design was modified from a striking reed pipe of an organ because he recorded that he wanted to reduce the rattle of his speech-imitating machine by modifying the reed. Furthermore, he did not mention the mouth organ at all in his work. An additional evidence for this hypothesis is his detailed figures of his speech imitating machine. They show clearly that the only difference between his free reed and the striking reed pipe of an organ is the smaller size of the reed in order to fit through the shallot. In contrast, Asian free reeds are flat, have no tuning wire, and the reed and reed opening are made from the same piece of metal.

2:50–3:00 Break

Invited Papers

3:00

2pMUa5. Laboratory measurements on free reeds from the reed organ, accordion, and khaen. Michael Busha (Phys. Dept., Grinnell College, Grinnell, IA 50112), James P. Cottingham, and Philip D. Koopman (Coe College, Cedar Rapids, IA 52402)

Several series of experimental measurements have been made on the motion of air-driven free reeds. These include individual reed organ reeds mounted on a laboratory wind chamber, accordion reeds mounted on the accordion reed block, and khaen reeds in individual bamboo pipes, in which the symmetric free reed is coupled to the pipe resonator. Measurements of reed displacement and velocity as a function of time have been made using a variable impedance transducer and a laser vibrometer system. Additional measurements made for each type of reed include sound pressure and air flow waveforms, as well as variation of sounding frequency and amplitude of vibration with blowing pressure. In comparing the sound pressure and air flow waveforms among the various types of reeds or for the same reed at different blowing pressures, differences can be understood in terms of the configuration of the reed and reed frame system and the amplitude of reed vibration. Some of the results can be interpreted in terms of simple theoretical models.

3:30

2pMUa6. From time domain simulation of diatonic harmonica, proposal of a minimal free reed model. Millot Laurent and Ambroise Daniel (Laboratoire d'Acoustique Musicale, 11 Rue de Lourmel, 75015 Paris, France)

Chromatical playing on diatonic harmonica can be explained and reproduced using a temporal model. This model features a nonlinear vocal track, two free reeds and interaction of these three elements. The agreement between numerical simulations and experiment will be illustrated. From these results, and considering the sinusoidal reed motion during normal playing, a minimal description of free reed will be proposed. The equivalent reed oscillator will be derived from comparisons between quasistatic and modal description of the reed displacement; it will be shown that the modal description is a good choice. Then, a sufficient description of the flow through each reed will be built. Waveforms and related sounds will be presented for each tested assumption and will illustrate the fact that this flow must rely on the local reed displacement. Application of this model will be discussed in the case of other free reed instruments.

4:00

2pMUa7. Acoustic coupling between oral tract and diatonic harmonica: Recent observations. James Antaki, Henry Bahnson, and Greg Burgreen (Dept. of Surgery and Bioengineering, Univ. of Pittsburgh, Pittsburgh, PA)

The diatonic harmonica is arguably one of the most ''vocal'' of all reed instruments. Due to the relatively short distance, thus tight coupling, between the player's oral cavity and the instrument, relatively subtle changes in the oral tract may effect dramatic modulations of both frequency and spectral content of the sound produced. Over the past several years we have endeavored to improve the understanding of this interaction, as well as the coupling between reeds of a given cell. This presentation will discuss our most recent observations. Benchtop experiments were conducted by laser vibrometer to measure the simultaneous motion of the reeds. Studies on human subjects were based on x-ray, ultrasonographic, and laryngoscopic imaging of players while sounding notes on the harmonica. Evidence of torsional instability has been observed, and analyzed by subsequent computational fluid dynamics simulation.

Session 2pMUb

Musical Acoustics: Free Reed Instruments: Concert Performance Featuring Henry Doktorski, Accordion and Howard Levy, Harmonica

James P. Cottingham, Cochair
Physics Department, Coe College, Cedar Rapids, Iowa 52402

James F. Antaki, Cochair
Department of Surgery and Bioengineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Chair's Introduction—5:00

This performance technical session will feature classical accordionist Henry Doktorski and harmonica virtuoso Howard Levy.

Henry Doktorski is one of the leading concert accordionists in the United States and a recognized figure in international classical accordion circles. He has performed on accordion with the Pittsburgh Symphony Orchestra, the Tanglewood Festival Orchestra, the Cleveland Chamber Symphony, the McKeesport Symphony and the Pittsburgh New Music Ensemble. Henry Doktorski serves on the faculty of The City Music Center at Duquesne University as instructor of accordion and is the founder of The Classical Free-Reed, Inc., a nonprofit educational organization devoted to the accordion and classical music. (See the Classical Free Reed web site at: http://trfn.clpgh.org/free-reed/)

Howard Levy is a musician without limits. His musical adventures include journeys into jazz, pop, rock, world music, Latin, classical, folk, blues, country, theater, and film. He has appeared on hundreds of CDs, won a Grammy, won a Joseph Jefferson Award for Best Original Music for a Play, and performed many times on American and European television. Universally acknowledged as the world's most advanced diatonic harmonica player, Howard has developed a fully chromatic style on the standard 10-hole diatonic harmonica, revolutionizing harmonica playing and taking the instrument into totally new territory. (Visit his web site at: http://www.levyland.com/)

Session 2pPA

Physical Acoustics and Underwater Acoustics: Recent Developments in Wave Propagation in Random and Complex Media II

D. Keith Wilson, Chair
AMSRL CI EE, U.S. Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland 20783-1197

Invited Papers

1:00

2pPA1. The effects of turbulent intermittency on scattering and estimates for the degree of saturation. David E. Norris (BBN Technologies, 1300 N. 17th St., Arlington, VA 22209, dnorris@bbn.com) and D. Keith Wilson (U.S. Army Res. Lab., Adelphi, MD 20783-1197)

Turbulent fluctuations in atmospheric wind and temperature fields are observed to be erratic in time; strong activity is typically interspersed with periods of relative calm. This property, referred to as turbulent intermittency, has an observable effect on the statistics of scattered acoustic signals. For fully saturated scattering, large intensity deviations about the mean result in a divergence from log-normal intensity probability density functions (pdfs) typically used to describe such statistics. Previous studies have developed the theory to predict the intensity pdf's that account for both turbulent intermittency and the degree of saturation [Norris *et al.*, J. Acoust. Soc. Am. **109**, 1871–1880 (2001); Wilson *et al.*, J. Acoust. Soc. Am. **99**, 3393–3400 (1996)]. The new formulation with intermittency is compared to the generalized gamma pdf previously proposed for propagation in random media. It is also compared to data collected over a 140 m line-of-sight path at 110 to 525 Hz. Experimental characterization of the pdf parameters relating to intermittency and signal saturation is attempted with partial success. Strength/diffraction parameters and complex signal moments are computed to further define the scattering effects as a function of frequency and help identify limitations in the theory.

1:20

2pPA2. Acoustic granular materials with pore size distribution close to log-normal. Kirill Horoshenkov and Mark Swift (Dept. of Civil and Environ. Eng., Univ. of Bradford, Bradford BD7 1DP, UK)

The water suction method is used to determine the parameters of the pore size distribution of a representative selection of loose and consolidated granular materials. It is shown that the experimentally determined pore size distribution in granular materials is often close to log-normal. The low- and high-frequency asymptotic behavior of the Biot viscosity correction function for media with log-normal pore size distribution is investigated and used to develop a simple rational approximation. This approximation is used to predict accurately the acoustic characteristic impedance and propagation constant for this class of materials. Unlike many available theoretical models for the acoustic properties of porous media which involve empirical shape factors, the proposed approximation is based entirely on four routinely measurable nonacoustic parameters: the porosity, flow resistivity, tortuosity, and the standard deviation of the pore size. The theoretical predictions for the acoustic surface impedance and absorption coefficient of loose and consolidated granulates are compared against the experimental results. A good agreement is obtained throughout the considered frequency range. [The authors are grateful to the Engineering and Physical Sciences Research Council (EPSRC) in the UK (Grant GR/L54905) for the support of this work. The authors would like to thank Dr. P. Leclaire for his help with the water suction experiments.]

1:40

2pPA3. Voronoi polygons and self-consistent technique used to compute the airflow resistivity of randomly placed fibers in glass wool. Viggo Tarnow (Dept. of Mech. Eng., Tech. Univ. of Denmark, Bygning 358, DK-2800 Kgs. Lyngby, Denmark)

Sound in glass wool propagates mainly in the air between glass fibers. For sound waves considered here, the distance between fibers is much smaller than the wavelength. Therefore, the sound velocity and attenuation can be computed from an effective mass density and compressibility. For simple harmonic waves at low frequencies, the effective mass density is determined by the friction between air and fibers. The friction is described by the airflow resistivity, which depends on frequency, but for frequencies below 1000 Hz in glass wool with density 15–30 kg/m^3, the resistivity to airflow is constant, and equal to the constant current value. A computation of resistivity from fiber density and diameter will be presented for a model of glass wool that consists of parallel randomly placed fibers with equal diameters. The computation is based on Voronoi polygons, and the results will be compared with results from a self-consistent technique.

2:00–2:15 Break

Contributed Papers

2:15

2pPA4. Scattering of acoustic evanescent waves by circular cylinders: Partial wave series solution. Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

Evanescent acoustical waves occur in a variety of situations such as when sound is incident on a fluid interface beyond the critical angle and when flexural waves on a plate are subsonic with respect to the surrounding fluid. The scattering by circular cylinders at normal incidence was calculated to give insight into the consequences on the scattering of the evanescence of the incident wave. To analyze the scattering, it is necessary to express the incident wave using a modified expansion involving cylindrical functions. For plane evanescent waves, the expansion becomes a double summation with products of modified and ordinary Bessel functions. The resulting modified series is found for the scattering by a fluid cylinder in an unbounded medium. The perfectly soft and rigid cases are also examined. Unlike the case of an ordinary incident wave, the counterpropagating partial waves of the same angular order have unequal magnitudes when the incident wave is evanescent. This is a consequence of the exponential dependence of the incident wave amplitude on the transverse coordinate. The associated exponential dependence of the scattering on the location of a scatterer was previously demonstrated [T. J. Matula and P. L. Marston, J. Acoust. Soc. Am. **93**, 1192–1195 (1993)].

2:30

2pPA5. A new numerical method for wave propagation through assemblies of cylinders and spheres. Takeru Yano (Dept. of Mech. Sci., Hokkaido Univ., Sapporo 060-8628, Japan) and Andrea Prosperetti (Johns Hopkins Univ., Baltimore, MD 21218)

PHYSALIS is a new method for the numerical solution of a variety of problems (potential theory, Navier-Stokes equations, and others) involving cylindrical or spherical internal boundaries [A. Prosperetti and H. N. Oguz, J. Comput. Phys. **167**, 196–216 (2001)]. At the heart of the method is the use of an exact analytical solution to transfer the boundary conditions from the surface of the inclusions to the neighboring grid nodes. This step avoids the difficulty deriving from the complex geometrical relationship between the internal boundaries and the underlying regular grid, with the added benefit that fast solvers can be used. In this work the method is adapted to two-dimensional acoustic scattering by cylinders as governed by the Helmholtz equation. As in prior applications, the method reveals itself highly efficient and of a relatively simple implementation. These features are illustrated on several problems. In particular, it is shown that the computational time grows much less than linearly with the number of cylinders, which permits the simulation of complex multiple scattering problems without large computational resources. [Work supported by The Japan Ministry of Education, Culture, Sports, Science and Technology, and by ONR.]

2:45

2pPA6. Acoustic wave propagation in atmospheric boundary layer above a rough terrain. Lanbo Liu and Mark L. Moran (U.S. Army Cold Regions Res. and Eng. Lab.)

In this paper we will present the latest development in the numerical simulation of acoustic wave propagation in the lowest atmospheric boundary layer above a rough terrain with a finite difference time domain (FDTD) algorithm. It is of great interest to investigate the effect of dynamic turbulent boundary layer meteorology coupled with surface terrain on acoustic wave propagation in frequency bands relevant to battlefield sensor systems. Turbulent mixing in the vicinity of the earth–air boundary layer is the key factor contributing to acoustic signal fluctuations such as upward refraction, scattering, high attenuation, and coherence degradations. Using first principles acoustic wave equation expressions for a moving fluid will allow us to couple acoustic models with time varying, 3-D large eddy simulation (LES) results. The specific meteorological parameters are vector wind velocity, temperature, pressure, and density as a function of space and time. Our acoustic simulation uses a coupled system of first-order temporal and spatial differential equations relating wind velocity, pressure, and density variations. This system of equations was evaluated with an explicitly time-stepping finite difference numerical ap-

proach. The coupling of LES and acoustic FDTD simulation will allow us to directly identify spectral portions of turbulent boundary layer meteorology contributing to short- and long-term acoustic signal variations.

3:00

2pPA7. The acoustics of plasma-sprayed, thick-thermal-barrier coatings. John Harris (Ctr. QEFP, Northwestern Univ., Evanston, IL 60208-3020)

Plasma-sprayed, thick-thermal-barrier coatings are very porous, approximately 2 mm thick, coatings used to insulate metal engine parts from high-temperature combustion processes. They are grown such that their elastic constants and density increase with depth; at the surface the elastic constants take small values, especially the shear modulus which is very small, but rapidly increase to those typical of the metal substrate. A Thomson–Haskell matrix technique is used to find numerically the eigenmode(s) along each vertical line through the coating and substrate, at each station along the direction of propagation. Using the lowest eigenmode, calculated at each station, a JWKB approximation to propagation in a coating with variable thickness is calculated, and from this the changes in travel time caused by varying degress of spalling at the surface are worked out. Using the two or three lowest eigenmodes for a coating of uniform thickness, a coupled-mode calculation is used to determine to what degree an interfacial inclusion couples the lowest eigenmode into the next higher one. In both cases the goal is to calculate the changes caused by the defects so that their presence may be detected. [Work supported by NSF.]

3:15

2pPA8. Performance of acoustic sensor arrays operating in atmospheric turbulence: Geometric acoustics regime. Sandra L. Collier and D. Keith Wilson (U.S. Army Res. Lab., AMSRL-CI-EE, 2800 Powder Mill Rd., Adelphi, MD 20783-1197)

The performance of acoustic sensor arrays operating in atmospheric turbulence with fluctuations described by a von Kármán spectrum has been investigated for strong diffraction in the presence of strong or weak scattering [Collier and Wilson, ASA June 2001 Meeting, ASA Dec. 2001 Meeting, J. Acoust. Soc. Am. (in submission)]. Here the previous analysis is extended to the geometric acoustics regime, i.e., to the weak diffraction and weak scattering limit. A model for the probability likelihood function is developed and the coherence function calculated. The Cramer–Rao lower bounds of the wavefront angles of arrival are investigated for geometric acoustics and are shown to reduce to the expected expression for the angle-of-arrival variance in the limit of zero noise.

3:30

2pPA9. An empirical model for the frequency spectrum of surface pressure fluctuations. Michael Goody (NSWC-CD, Code 7250, 9500 MacArthur Blvd., West Bethesda, MD 20817-5700, GoodyMC@nswccd.navy.mil)

An empirical model is presented for the surface pressure fluctuations beneath a two-dimensional, zero-pressure-gradient, turbulent boundary layer. The model is based on the form of Chase's model [J. Sound Vib. **70** (1), 29–67 (1980)] given by Howe (*Acoustics of Fluid-Structure Interactions* (Cambridge U.P., Cambridge, 1998), p. 208]. The model of Chase is adjusted to better agree with several well-respected experimental data sets that cover the momentum Reynolds number range, $1400 < \mathrm{Re}_\theta < 23\,400$. The new model requires knowledge *a priori* of the free-stream velocity, friction velocity, displacement thickness, and kinematic viscosity. The new model also incorporates the observed scaling behavior of the pressure spectrum. The size of the midfrequency overlap region of the model spectrum is Reynolds-number dependent. Additionally, the model spectrum beneath flows with different Reynolds numbers collapse to a single curve that decays as ω^{-5} at high frequencies when normalized using viscous scales. By collapsing at high frequencies when normalized on viscous scales, the model offers a high degree of confidence when extrapolated to flows with a higher Reynolds number ($\mathrm{Re}_\theta > 23\,400$). Examples are given using analytically calculated flows that extend to $\mathrm{Re}_\theta = 2.45 \times 10^6$. Integration of the model equation recovers the measured p'/τ_W to within 8%.

TUESDAY AFTERNOON, 4 JUNE 2002 GRAND BALLROOM 4, 2:00 TO 3:45 P.M.

Session 2pPP

Psychological and Physiological Acoustics: Hearing Aid Design: Psychophysics and Signal Processing II

Dianne J. Van Tasell, Cochair
Starkey Laboratories, 6600 Washington Avenue, South, Eden Prairie, Minnesota 55344

Brent W. Edwards, Cochair
Sound ID, 3430 West Bayshore Road, Palo Alto, California 94303

Contributed Papers

2:00

2pPP1. Fast dynamic range compression method for multichannel hearing aids. David V. Anderson (Georgia Inst. of Technol., Atlanta, GA 30332)

Hearing loss with recruitment results in a frequency-dependent reduced effective dynamic range of the ear. The advance of digital hearing aid capabilities has led to the proliferation of multiband, compression algorithms directed at the recruitment problem. However, results have been mixed at best. We propose an alternative design methodology to designing multiband compressive aids based on attack and release times. In particular, multiband compression hearing aids should be matched to certain characteristics of the ear with a memoryless gain based on the bandlimited envelope function of each band. Under the proposed design method each band has a different effective time-constant.

2:15

2pPP2. A scheme for noise suppression and spectral enhancement of speech to alleviate speech reception problems from loss of frequency selectivity. Johannes Lyzenga, Joost M. Festen, and Tammo Houtgast (Dept. of ENT/Audiol., VU Medical Ctr., Amsterdam, The Netherlands)

Even after sufficient amplification, hearing-impaired listeners often experience problems in understanding speech under noisy conditions. This may be caused by suprathreshold deficits such as loss of compression and reduced frequency selectivity. In this project we investigate a scheme in which speech and noise are processed before presentation to try and alleviate intelligibility problems caused by reduced frequency selectivity. The scheme contains three strategies, one in which the peak-to-valley ratios of selected modulations in the speech spectrum are enlarged, a second in which the overall speech spectrum is modified, and a third in which noise is suppressed before the two enhancement steps. An overlap-and-add (OLA) algorithm is used in the implementation. The effect of the speech processing is evaluated by measuring speech-reception thresholds (SRT) for sentences in speech noise, estimating the signal-to-noise ratio at which listeners can correctly reproduce 50% of presented sentences. Hearing-impaired and normal-hearing listeners were used. To simulate the hearing impairment resulting from a loss of frequency selectivity, we spectrally smeared the stimuli presented to the normal-hearing listeners. We found that the preprocessing scheme achieved a modest improvement of nearly 2 dB in the SRT for normal-hearing listeners. Data for hearing-impaired listeners are presently being collected.

2:30

2pPP3. Directional hearing aid using hybrid adaptive beamformer (HAB) and binaural ITE array. Scott T. Shaw, Andy J. LaRow, Gary L. Gibian (Planning Systems, Inc., 12030 Sunrise Valley Dr., Reston, VA 20191, ggibian@plansys.com), LaGuinn P. Sherlock (Univ. of Maryland School of Medicine, Baltimore, MD 21201), and Robert Schulein (Etymotic Res., Elk Grove Village, IL 60007)

A directional hearing aid algorithm called the Hybrid Adaptive Beamformer (HAB), developed for NIH/NIA, can be applied to many different microphone array configurations. In this project the HAB algorithm was applied to a new array employing in-the-ear microphones at each ear (HAB-ITE), to see if previous HAB performance could be achieved with a more cosmetically acceptable package. With diotic output, the average benefit in threshold SNR was 10.9 dB for three HoH and 11.7 dB for five normal-hearing subjects. These results are slightly better than previous results of equivalent tests with a 3-in. array. With an innovative binaural fitting, a small benefit beyond that provided by diotic adaptive beamforming was observed: 12.5 dB for HoH and 13.3 dB for normal-hearing subjects, a 1.6 dB improvement over the diotic presentation. Subjectively, the binaural fitting preserved binaural hearing abilities, giving the user a sense of space, and providing left–right localization. Thus the goal of creating an adaptive beamformer that simultaneously provides excellent noise reduction and binaural hearing was achieved. Further work remains before the HAB-ITE can be incorporated into a real product, optimizing binaural adaptive beamforming, and integrating the concept with other technologies to produce a viable product prototype. [Work supported by NIH/NIDCD.]

2:45

2pPP4. Recruitment-of-loudness effects of attenuative noise reduction algorithms. Nathaniel Whitmal III (ECE Dept., Worcester Polytechnic Inst., Worcester, MA 01002) and Azadeh Vosoughi (Cornell Univ., Ithaca, NY 14853)

Hearing-impaired listeners have greater difficulty understanding speech in noise than normal-hearing listeners do. As a result, hearing aid users are often challenged by the inability of their hearing aids to improve intelligibility in noise. Several investigators have addressed this problem by using well-known signal processing methods (e.g., spectral subtraction, Wiener filtering) to enhance noise-corrupted speech. Unfortunately, these methods have failed to provide significant improvements in intelligibility. One possible explanation is the level-dependent nature of the attenuation that the algorithms impose on the speech. In the cases described above, this attenuation resembles the piecewise-linear input–output characteristic observed in certain recruitment-of-loudness simulators. The purpose of this study was to compare the intelligibility of processed speech with that expected for recruitment-of-loudness simulation. Trials of the CUNY Nonsense Syllable Test were conducted with 12 normal-hearing listeners, using syllables that were mixed with additive noise at SNRs of 6, 12, and 18 dB. Input–output characteristics for the signals were measured and used to determine the effective threshold shift imposed by the algorithms. Comparisons of measured intelligibility scores with articulation index-based intelligibility predictions indicate that the behavior of such noise reduction algorithms can be successfully modeled as a form of mild sensorineural hearing loss.

3:00

2pPP5. Optimization of speech in noise with three signal processing algorithms for normal-hearing and hearing-impaired subjects. Bas A. M. Franck, Wouter A. Dreschler (Academic Medical Ctr., Dept. of Clinical and Exp. Audiol., D2-223, P.O. Box 1100 DD, Amsterdam, The Netherlands), and Johannes Lyzenga (Free Univ. Medical Ctr., Amsterdam, The Netherlands)

In this study a three-dimensional Simplex procedure was applied to optimize speech in noise by a combination of signal processing algorithms for different acoustic conditions and hearing losses. The algorithms used to span the three dimensions are noise reduction, spectral tilting, and spectral enhancement, respectively. Additionally, we studied the algorithms for their main effects and interaction effects within the optimization process. The subjects were asked to evaluate two consecutive, differently processed sentences on listening comfort. Three different noise types and two signal-to-noise ratios (S/N) were used. Three groups of subjects participated: normal hearing, normal hearing with simulated impaired auditory filtering (by spectral smearing), and sensorineurally hearing-impaired subjects. For the normal-hearing group we applied $S/N=0$ dB. For the hearing-impaired and the simulated hearing-impaired subjects we applied $S/N=5$ dB. We will discuss the similarities and differences in the response patterns of the three groups. Also, the individual preferences will be related to the hearing capacity, and to the type of interfering noise. Finally, we will discuss differences in the perceptual features that are used to judge listening comfort of the fragments by normal-hearing and hearing-impaired subjects.

3:15

2pPP6. Physical modeling of the feedback path in hearing aids with application to adaptive feedback cancellation. Joanna L Hayes and Boaz Rafaely (Inst. of Sound and Vib. Res., Univ. of Southampton, Southampton SO17 1BJ, UK)

Hearing aid system modeling based on two-port network theory has been used previously to study the forward gain and the feedback path in hearing aids. The two-port modeling approach is employed in this work to develop an analytic model of the feedback path by reducing the model matrices to simplified analytic expressions. Such an analytic model can simulate the frequency response of the feedback path given the values of relatively few physical parameters such as vent dimensions. The model was extended to include variability in the feedback path due to slit leaks, for example. The analytic model was then incorporated in an adaptive feedback cancellation system, where the physical parameters of the model were adapted to match the actual feedback path and cancel the feedback signal. In the initial stage of this study, the ability of the model to match the frequency response of various measured feedback paths was studied using numerical optimization. Then, an adaptive filtering configuration based on the physical model was developed and studied using computer simulations. Results show that this new approach to adaptive feedback cancellation has the potential to improve both adaptation speed and performance robustness.

3:30

2pPP7. A novel method of improving sound quality and reducing acoustic feedback in hearing aids. Mead Killion, John French, Steve Viranyi, and David Preves (Etymotic Res., 61 Martin Ln., Elk Grove Village, IL 60007)

Most current hearing aids have relatively narrow bandwidths, when compared to high-fidelity equipment, and exhibit undamped peaks because the peaks are considered less troublesome than the problem of wax-clogged dampers. Attempting to make hearing aids wider band has typically resulted in increased acoustic feedback problems. The recent availability of an off-the-shelf digital hearing aid integrated circuit amplifier, which contains several biquad filters, when used with special software, automatically detects and suppresses peaks. The filters then further flatten and extend the hearing aid frequency response to 16 kHz, while the appropriate CORFIG correction is added to the frequency response, producing a transparent sound. Open ear versus aided KEMAR recordings were produced using a live jazz trio and a string quartet. The sound quality ratings for eight commercially available digital hearing aids were obtained from several different listening panels. The new response equalization proved advantageous in all cases. The effects of eliminating the peaks in the response on maximum real ear gain achievable before onset of acoustic feedback oscillation will be reported.

Session 2pSA

Structural Acoustics and Vibration: Pyroshock II

Harry Himelblau, Chair

Space & Communication Division, The Boeing Company, Rocketdyne Propulsion and Power, Canoga Park, California 91309-7922

Invited Papers

1:00

2pSA1. Statistical analysis of pyroshock data. William O. Hughes (NASA Glenn Res. Ctr., M/S 86-10, Cleveland, OH 44135-3191, william.o.hughes@grc.nasa.gov)

The sample size of aerospace pyroshock test data is typically small. This often forces the engineer to make assumptions on its population distribution and to use conservative margins or methodologies in determining shock specifications. For example, the maximum expected environment is often derived by adding 3–6 dB to the maximum envelope of a limited amount of shock data. The recent availability of a large amount of pyroshock test data has allowed a rare statistical analysis to be performed. Findings and procedures from this analysis will be explained, including information on population distributions, procedures to properly combine families of test data, and methods of deriving appropriate shock specifications for a multipoint shock source.

1:30

2pSA2. Pyroshock testing—shock simulation facilities. Vesta I. Bateman (Sandia Natl. Labs., P.O. Box 5800, Albuquerque, NM 87185-0555, vibatem@sandia.gov)

A variety of shock simulation facilities are available to simulate pyroshock events. These facilities range from bounded impact shock machines and electrodynamic shakers to resonant fixture techniques. This presentation will focus on the use of general purpose and tuned resonant fixture techniques including a unique tunable beam apparatus developed at SNL. Examples of application of the resonant fixture technique for both component and full-scale structure pyroshock simulations will be presented. Advantages and disadvantages of each technique will be discussed along with the usable frequency content and bandwidth.

2:00

2pSA3. Pyroshock testing—electrodynamic shakers. David O. Smallwood (Sandia National Labs., P.O. Box 5800, Albuquerque, NM 87185-0553, dosmall@sandia.gov)

Far field pyroshock (accelerations less than a few hundred grams, and bandwidths less than a few kHz) can be simulated on electrodynamic shakers. Typically, the specification is in terms of the shock response spectrum (SRS). Wave forms are synthesized which will match the required SRS. The process is not unique, as many wave forms can have essentially the same SRS. Sometimes additional restrictions are placed on the synthesized wave form. Most common are restrictions on the duration of the wave form. The process of synthesizing wave forms, which will match an SRS and conform to the limitations of electrodynamic shakers, will be described. The methods used to reproduce these wave forms on the shaker will then be discussed.

2:30–2:45 Break

2:45

2pSA4. Pyroshock testing: Effects of equipment simulators. William O. Hughes (NASA Glenn Res. Ctr., M/S 86-10, Cleveland, OH 44135-3191, william.o.hughes@grc.nasa.gov)

It is critical to a spacecraft's mission success that flight equipment be properly qualified to its expected shock environment. Simulators of varying fidelity are often used to measure the shock levels during spacecraft shock tests. Pyroshock test data from a recent spacecraft separation test will be shown that illustrates how shock response spectra differ for avionics equipment simulators of both low and high fidelity. The effect of the simulator weight on the shock response will also be shown. Additionally, data showing the attenuation of pyroshock levels, including its reduction with distance and across joints will be discussed. This observed attenuation data generally supports the standard methodologies for predictions of shock attenuation.

3:15

2pSA5. Pyroshock design. Allan G. Piersol (Piersol Eng. Co., 23021 Brenford St., Woodland Hills, CA 91364-4830, apiersol@pacbell.net)

The structural response velocities produced by pyroshocks are usually too low to cause structural damage beyond the immediate region of the pyrotechnic device where damage is intended. On the other hand, pyroshocks can cause malfunctions of equipment mounted near the pyrotechnic device, particularly equipment containing microelectronic or ceramic elements. These malfunctions are usually the result of the high frequency content (above 2 kHz) of pyroshocks, and are difficult to design against using conventional analysis procedures. On the other hand, adherence to simple design guidelines, which are summarized, can substantially suppress pyroshock-induced failures of equipment.

3:45–5:00
Panel Discussion

TUESDAY AFTERNOON, 4 JUNE 2002 GRAND BALLROOM 2, 1:00 TO 5:00 P.M.

Session 2pSC

Speech Communication: Models of Phonetic Category Formation and Structure

Lori L. Holt, Cochair
Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

Andrew J. Lotto, Cochair
Department of Psychology, Washington State University, P.O. Box 644820, Pullman, Washington 99164-4820

Chair's Introduction—1:00

Invited Papers

1:05

2pSC1. Speech perception as complex auditory categorization. Lori L. Holt (Dept. of Psych. and Ctr. for the Neural Basis of Cognition, Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213, lholt@andrew.cmu.edu)

Despite a long and rich history of categorization research in cognitive psychology, very little work has addressed the issue of complex auditory category formation. This is especially unfortunate because the general underlying cognitive and perceptual mechanisms that guide auditory category formation are of great importance to understanding speech perception. I will discuss a new methodological approach to examining complex auditory category formation that specifically addresses issues relevant to speech perception. This approach utilizes novel nonspeech sound stimuli to gain full experimental control over listeners' history of experience. As such, the course of learning is readily measurable. Results from this methodology indicate that the structure and formation of auditory categories are a function of the statistical input distributions of sound that listeners hear, aspects of the operating characteristics of the auditory system, and characteristics of the perceptual categorization system. These results have important implications for phonetic acquisition and speech perception.

1:30

2pSC2. Neuroimaging studies of phonetic category learning. Julie Fiez (Dept. of Psych. and Ctr. for the Neural Basis of Cognition, Univ. of Pittsburgh, 605 LRDC, Pittsburgh, PA 15260, fiez+@pitt.edu)

The Japanese perception of /r/ vs /l/ is being used as a model system to explore a Hebbian-based model of adaptive plasticity. Prior behavioral work has demonstrated the effectiveness of an adaptive training regime and the utility of response feedback on the acquisition of the ability to discriminate a minimal speech pair (e.g., ''rock'' vs ''lock''). Imaging, using fMRI methods, is now being used to explore the neural substrates that underlie changes in categorical perception. Two questions are being addressed: (1) in which brain areas can a differential response to /r/ vs /l/ be identified, and is activity in these regions influenced by training? and (2) can the effectiveness of feedback as a part of the training regime be explained by the recruitment of reward-related brain regions that support learning?

1:55

2pSC3. Effects of category learning on auditory perception and cortical maps. Frank H. Guenther (Dept. of Cognit. and Neural Systems, Boston Univ., 677 Beacon St., Boston, MA 02215, guenther@cns.bu.edu)

Our ability to discriminate sounds is not uniform throughout acoustic space. One example of auditory space warping, termed the perceptual magnet effect by Kuhl and colleagues, appears to arise from exposure to the phonemes of an infant's native language. We have developed a neural model that accounts for the magnet effect in terms of neural map dynamics in auditory cortex. This model predicts that it should be possible to induce a magnet effect for non-speech stimuli. This prediction was verified by a psychophysical experiment in which subjects underwent categorization training involving non-speech auditory stimuli that were not categorical prior to training. The model further predicts that the magnet effect arises because prototypical vowels have a smaller cortical representation than non-prototypical vowels. This prediction was supported by an fMRI experiment involving prototypical and non-prototypical examples of the vowel /i/. Finally, the model predicts that categorization training with non-speech stimuli should lead to a decreased cortical representation for stimuli near the center of the category. This prediction was supported by an fMRI experiment involving categorization training with non-speech auditory stimuli. These results provide strong support for the model's account of the effects of category learning on auditory perception and auditory cortical maps.

2:20

2pSC4. On the flexibility of phonetic categories. Alexander L. Francis (Speech and Hearing Sci., Univ. of Hong Kong, Prince Philip Dental Hospital 5F, 34 Hospital Rd) and Howard C. Nusbaum (Univ. of Chicago, Chicago, IL)

Research on phonetic category acquisition suggests that over the course of development children shift the relative weight given to various acoustic cues to particular phonetic categories [S. Nittrouer, J. Phon. **20**, 1–32 (1992)]. Adults also show flexibility in acquiring second language categories and learning to recognize synthetic speech. In both cases listeners must shift attention among some acoustic cues and change the weighting and mapping of others. Recent research demonstrated that category-level feedback can induce adult listeners to shift attention between conflicting cues to native stop-consonant categories. Subsequent research showed that adult listeners also reorganize the distribution of their attention to acoustic cues in learning to better understand synthetic speech. Moreover, research on learning foreign phonetic contrasts suggests that listeners can learn to direct attention to unfamiliar acoustic cues. These results suggest that phonetic categories are flexible in terms of the nature and quality of the acoustic cues that define them. We argue that this flexibility is a consequence of the basic nature of the general mechanism that mediates speech perception: In order to resolve the problem of context-conditioned variability, listeners must dynamically modify attention to the acoustic signal in a context-dependent manner.

2:45–3:00 Break

3:00

2pSC5. Perceptual interference effects on phonetic categorization by second-language learners and cochlear implant patients. Paul Iverson (Dept. of Phonet. and Linguist., Univ. College London, 4 Stephenson Way, London NW1 2HE, UK)

This talk will outline a new theory that describes how distortions in auditory processing, due to language experience or to hearing impairment, can interfere with phonetic categorization processes. Experimental data will be presented on the perception of American English /r/ and /l/ by American and Japanese listeners. Native-language tests of adult cochlear implant patients, for the /t/–/d/ contrast, will also be presented. The results suggest that the formation and adaptability of phonetic categories are impaired when the auditory processing of speech stimuli increases the salience of within-category acoustic variation.

3:25

2pSC6. Internal structure of phonetic categories: Some characteristics and constraints. Joanne L. Miller (Dept. of Psych., Northeastern Univ., Boston, MA 02115)

A widely held assumption in the speech perception literature for many years was that during the course of processing listeners derive an abstract phonetic representation and, in doing so, discard information about the fine-grained detail of the speech signal. However, more recent research has shown that the representations of speech are much richer than this emphasis on abstract categories would suggest, and that listeners retain in memory a substantial amount of fine-grained acoustic-phonetic information. One line of

evidence for the richness of phonetic representations comes from research showing that phonetic categories are internally structured in a graded fashion, with some members of the category perceived as better exemplars (as more ''prototypical'') than others. In this talk I will describe findings from our research program that highlight some of the characteristics of these internally structured categories, and discuss how these characteristics place constraints on models of phonetic perception. [Work supported by NIDCD.]

3:50

2pSC7. Animal models of speech sound categorization. Andrew J. Lotto (Dept. of Psych., Washington State Univ., Pullman, WA 99164-4802, alotto@wsu.edu)

Several recent studies have examined phonetic category formation by training nonhuman animals (e.g., birds) to respond differentially to distributions of speech sounds. Animal models provide good tests of theories of perceptual category formation because one can completely control the input distributions and obtain a fine-grained sampling of responses to the stimulus space. In studies using vowel sound distributions, animals' response structures reveal two salient characteristics. First, there is a tendency to respond most strongly to stimuli that are furthest from the boundaries between training distributions. Second, there is evidence that the structure of responses is a function of the statistics of the training input. This can be seen as a peak in response for stimuli near the centroid of the input distribution. These same response structure characteristics can also be seen in data collected from humans in categorization tasks. It is possible that these two characteristics are indicative of two systems of learning: one functional and one statistical. A similar dichotomy has also been recently proposed in the visual categorization literature. The implications of these results for theoretical and computational models of phonetic acquisition will be discussed. [Work supported by NSF and NIH.]

Contributed Papers

4:15

2pSC8. The interaction of short-term and long-term memory in phonetic category formation. James D. Harnsberger (Dept. of Commun. Sci. & Disord., Univ. of Florida, 336 Dauer Hall, P.O. Box 11742, Gainesville, FL 32611-7420)

This study examined the role that short-term memory capacity plays in the relationship between novel stimuli (e.g., non-native speech sounds, native nonsense words) and phonetic categories in long-term memory. Thirty native speakers of American English were administered five tests: categorial AXB discrimination using nasal consonants from Malayalam; categorial identification, also using Malayalam nasals, which measured the influence of phonetic categories in long-term memory; digit span; nonword span, a short-term memory measure mediated by phonetic categories in long-term memory; and paired-associate word learning (word–word and word–nonword pairs). The results showed that almost all measures were significantly correlated with one another. The strongest predictor for the discrimination and word–nonword learning results was nonword ($r = +0.62$) and digit span ($r = +0.51$), respectively. When the identification test results were partialed out, only nonword span significantly correlated with discrimination. The results show a strong influence of short-term memory capacity on the encoding of phonetic detail within phonetic categories and suggest that long-term memory representations regulate the capacity of short-term memory to preserve information for subsequent encoding. The results of this study will also be discussed with regards to resolving the tension between episodic and abstract models of phonetic category structure.

4:30

2pSC9. Formation of categories for complex novel auditory stimuli. Daniel Mirman, Lori L. Holt, and James L. McClelland (Ctr. for the Neural Basis of Cognition and Psych. Dept., Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213, dmirman@andrew.cmu.edu)

Categorization of complex sounds with multiple, imperfectly valid cues is fundamental to phonetic perception. To study the general perceptual and cognitive processes that support complex sound categories, a novel stimulus set was created that allows tight control of category structure and input distributions. Stimuli were created from 300-ms noise bursts by applying bandstop filters at varying center frequencies and manipulating rise/fall time of stimulus onset and offset. Stimuli were assigned to one of two categories and presented to participants in a category identification and an AX discrimination task. Feedback was provided during identification trials, but not during discrimination tasks. Participants quickly learned to apply the category labels with high accuracy. Identification reaction times followed a pattern typical of speech stimuli with an apex in reaction time at category boundary. These results are consistent with formation of new auditory categories. Preliminary results indicate that discrimination performance is not tightly coupled with development of sharp identification functions and response-time peaks at category boundaries. Implications for mechanisms of speech categorization and category formation will be discussed. [Work supported by CNBC, NIH, and NSF.]

4:45–5:00
Panel Discussion
Randy Diehl, Moderator

Session 2pSP

Signal Processing in Acoustics: Signal Processing Techniques

David C. Swanson, Chair
Applied Research Laboratory, Pennsylvania State University, State College, Pennsylvania 16804

Contributed Papers

1:15

2pSP1. Forward-scattered acoustic intensity from prolate spheroids. Brian R. Rapids and Gerald C. Lauchle (Grad. Prog. in Acoust. and Appl. Res. Lab., Pennsylvania State Univ., P.O. Box 30, State College, PA 16804)

In bistatic scattering geometries, the detection of a forward-scattered signal is particularly difficult because the incident and scattered waves combine into a simultaneous mixture. The result is that the amplitude of the scattered wave becomes masked by that of the incident wave. During detection of the scattered signal, conventional space-time processing techniques regard the source signal as interference and attempt to suppress it. While vector sensors alone possess an inherent directivity due to their fundamental nature, intensity vector sensors coherently measure the acoustic pressure and particle velocity components (or related quantity such as acceleration, displacement, or pressure gradient). The coherent measurement of both acoustic field parameters may provide unconventional information regarding the presence of an object because of their known relationship. It is hypothesized that techniques based upon these coherent measurements will be able to process the total acoustic field rather than filter the scattered signal from the incident signal during detection. Theoretical and available experimental results will be presented to describe these hypothesized capabilities. [Work supported by ONR, Code 321SS under Grant No. N00014-01-1-0108, Dr. James F. McEachern, project monitor.]

1:30

2pSP2. Further investigations into performance metrics for underwater communications. Scott L. Whitney, Geoffrey S. Edelson, Ned B. Thammakhoune, and Michael S. Richman (BAE Systems, MER15-2651, P.O. Box 868, Nashua, NH 03061-0868, scott.l.whitney@baesystems.com)

Determining the relationship between the performance of an underwater acoustic data communications system and the operating environmental conditions is a problem that continues to plague researchers. The complexity of the time-varying channel is difficult to measure and model. Therefore an approach that uses metrics measured from data collected at sea to characterize the channel is attractive. As expected, preliminary assessments on limited data have shown that performance depends not only on environmental conditions, but also on system implementation. By extracting a variety of metrics, a better understanding of the subset that discriminate between good and bad performance can be developed. Also, by analyzing the relationship between certain metrics and performance, system limitations can be identified for re-evaluation. For example, a surprising result of the initial assessment of performance using a multichannel decision feedback equalizer on real data showed that sparseness of multipath arrivals may be an arbiter of performance [M. S. Richman *et al.*, J. Acoust. Soc. Am. **110**, 2619 (2001)]. Therefore changes to the algorithm that allows for sparse arrivals may improve performance. In this paper, a larger number of metrics from greater quantities of real-data and system configurations are measured and evaluated against equalizer results.

1:45

2pSP3. Reverberation noise modeling using extreme value theory. Brian La Cour and Robert Luter (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, blacour@arlut.utexas.edu)

Normalized matched filter output forms the basis of target detection in active sonar. In a target-free environment, the central theorem, if valid, predicts that the statistics of the envelope follow a Rayleigh distribution, and, to first approximation, this is indeed observed. However, well-known departures from the Rayleigh model are found in the tail end of observed distributions. Traditional approaches to this problem have focused on constructing a simple, parameterized, non-Rayleigh distribution which more closely models observations. This paper suggests a novel alternative which focuses on a robust method of modeling only the tails of the distribution in favor of the less important body. Results from extreme-value theory are used to fit a generalized Pareto distribution (GPD) to the empirical cumulative distribution function, conditioned on a large threshold value. [A random variable X has a GPD if $P(X \leqslant x) = 1 - (1 + \gamma x/\sigma)^{-1/\gamma}$ for $x \geqslant 0$, $\sigma > 0$, and γ real; $\gamma = 0$ is the exponential distribution.] Estimates of γ and σ are discussed for a broad range of active sonar data, and the results are compared with fits to other popular non-Rayleigh models. The origins of non-Rayleighness are also considered, including finite-size effects, spatial and temporal correlations, and nonuniformity.

2:00

2pSP4. Performance of conventional and fluctuation-based signal detection applied to atmospheric acoustics in the presence of transients. Thomas Null, Chris Clark (Mil Tec, NCPA, Coliseum Dr., University, MS 38677), and R. A. Wagstaff (Univ. of Mississippi, University, MS 38677)

One problem encountered in atmospherics acoustics is the detection of steady signals in the presence of loud transient noise. The ability to discriminate against loud transients is an attractive feature of fluctuation-based beamforming. A fluctuation-based beamformer was developed and there was a need to evaluate its performance for acoustic environments that have frequently occurring transients. Synthetic noise fields, which included loud transient noise, were created. Subsequently, receiver operating characteristic (ROC) curves could be produced in a Monte Carlo fashion. In order to provide a benchmark, the results of a conventional beamformer were similarly tested via ROC curves. These ROC curves allowed for comparison of the two beamformers under specific signal and noise conditions. In this presentation, the effects of transients on the outputs of both the conventional beamformer and the fluctuation-based beamformer are discussed. Particular interest is focused on the amplitude distribution of the outputs.

2:15

2pSP5. Construction of high frame rate images with Fourier transform. Hu Peng and Jian-yu Lu (Ultrasound Lab., Dept. of Bioengineering, The Univ. of Toledo, Toledo, OH 43606, jilu@eng.utoledo.edu)

Traditionally, images are constructed with a delay-and-sum method that adjusts the phases of received signals (echoes) scattered from the same point in space so that they are summed in phase. Recently, the

relationship between the delay-and-sum method and the Fourier transform is investigated [Jian-yu Lu, Anjun Liu, and Hu Peng, ''High frame rate and delay-and-sum imaging methods,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control (submitted)]. In this study, a generic Fourier transform method is developed. Two-dimensional (2-D) or three-dimensional (3-D) high frame rate images can be constructed using the Fourier transform with a single transmission of an ultrasound pulse from an array as long as the transmission field of the array is known. To verify our theory, computer simulations have been performed with a linear array, a 2-D array, a convex curved array, and a spherical 2-D array. The simulation results are consistent with our theory. [Work supported in part by Grant 5RO1 HL60301 from NIH.]

2:30–2:45 Break

2:45

2pSP6. Using cross-frequency cost functions for broadband source localization and environmental inversion. Ethan P. Honda (Appl. Res. Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758)

Cost functions that are constructed by coherently summing (model-to-data) correlations over hydrophone pairs and frequency have been used successfully for source localization [E. K. Westwood, J. Acoust. Soc. Am. **91**, 2777–2789 (1992)] as well as source localization and environmental inversion [Neilsen, J. Acoust. Soc. Am. (to be published)]. Although the coherent sum is usually taken over the same frequency for both data and model, it is shown that summing over other regions of the $f_{\text{data}} \otimes f_{\text{model}}$ space is also useful and may facilitate more efficient source localization. It is shown that lines of constant $f_{\text{data}}/f_{\text{model}}$ correspond to different source bearings. Although looking along lines of constant $f_{\text{data}}/f_{\text{model}}$ can be used as a crude form of spatial filtering, a new non-plane-wave spatial filter is also constructed that helps localize sources in the presence of multiple interferers. The spatial filter employed uses the environmental model to construct its set of basis functions and is therefore theoretically capable of spatially filtering in full 3-D as opposed to just bearing, as is done in adaptive beamforming.

3:00

2pSP7. Wigner–Ville representations for acoustic source localization. Zoi-Heleni Michalopoulou (Dept. of Mathematical Sci., New Jersey Inst. of Technol., Newark, NJ 07102) and Leon Cohen (Dept. of Phys. and Astron., Hunter College, City Univ. of New York, 695 Park Ave., New York, NY 10021)

Signal dispersion in a waveguide can be linked to source and receiver location as well as physical properties of the propagation medium. Traditionally, the main methods used for source localization in the ocean and in geoacoustic inversion problems have been based on the use of the spectrogram for the extraction of the dispersion information. In this work we explore the possibility of using other time-frequency transforms because other transforms, such as the Wigner–Ville representation, reflect the dispersion properties of the waveguide more accurately than conventional spectrograms. We apply the Wigner–Ville distribution, in conjunction with sound propagation models, for inversion with underwater sound. Results with synthetic data calculated for simplified ocean media indicate the potential of the approach for successful parameter estimation. [Work supported by ONR.]

3:15

2pSP8. Multichannel active noise control and acoustic equalization using fast affine projection algorithms. Martin Bouchard (School of Information Technol. and Eng., Univ. of Ottawa, 161 Louis Pasteur, Ottawa, ON K1N 6N5, Canada)

In the field of adaptive signal processing, it is well known that affine projection algorithms or their low-computational implementations, fast affine projection algorithms, can produce a good trade-off between convergence speed and computational complexity. Although these algorithms typically do not provide the same convergence speed as recursive-least-squares algorithms, they can provide a much improved convergence speed compared to stochastic gradient descent algorithms, without the high increase of the computational load or the instability often found in recursive-least-squares algorithms. In this presentation, multichannel fast affine projection algorithms are introduced for active noise control or acoustic equalization. Multichannel fast affine projection algorithms have been previously published for acoustic echo cancellation, but the problem of active noise control or acoustic equalization is a very different one, leading to different structures. The computational complexity of the new proposed algorithm is evaluated, and it is shown through simulations that not only can the new algorithm provide the expected trade-off between convergence performance and computational complexity, it can also provide the best convergence performance (even over recursive-least-squares algorithms) when non-ideal noisy acoustic plant models are used in the adaptive systems.

3:30

2pSP9. A comparison of algorithms and the development of a new fast convergence and reduced computational load algorithm for multichannel active noise control. Martin Bouchard and Scott Norcross (School of Information Technol. and Eng., Univ. of Ottawa, ON K1N 6N5, Canada)

In this presentation, the three main factors that affect the convergence speed of learning algorithms for adaptive FIR filters used in multichannel active noise control are described. Based on these three factors, a comparison of several adaptive FIR filter algorithms for multichannel active noise control is done, including several existing algorithms and a few unpublished algorithms. Of the unpublished algorithms, one algorithm has the potential for optimal convergence speed, and this algorithm is described in more detail in the presentation. The algorithm combines the use of recursive-least-squares algorithms with the use of an inverse model of the multichannel acoustic plant between the actuators and the error sensors. The resulting algorithm is called the multichannel inverse delay-compensated filtered-x RLS algorithm for active noise control. This algorithm can not only provide fast convergence, but for multichannel systems it also provides a significant reduction of the computational load compared to the previously published algorithm with the fastest convergence speed. Simulation results are presented to validate the convergence behavior of the new proposed algorithm.

3:45

2pSP10. Inverse source problem by convex optimization with constraints over the object space and signal field. Kenbu Teramoto (Dept. of Mech. Eng., Saga Univ., Saga-shi 8408502, Japan, tera@me.saga-u.ac.jp)

In the acoustical endscopy, due to the physical limitations, the transducer array is composed of a small number of elements and each interspacing is larger than the acoustical wavelength that is called a sparse array system. In such cases, avoiding the ill-posed problems, projection onto convex sets (POCS) methods are used with incorporating constraints about both the signal field and the object space. POCS, however, is based on the alternating projections paradigm, which has a slow-convergence property in general. Furthermore if inconsistency exists in the set of constraints, this POCS algorithm cannot guarantee the convergence to the optimal estimate. The proposed algorithm is based on convex optimization over the direct product of the object space and the observed signal field. By acoustical experiments, it is proved that the proposed algorithm has the following improvements: (1) Targets can be identified when unknown components exist in the transfer function. (2) Transient behavior of the convergence becomes more stable than that of POCS algorithm. (3) Instability caused by the inconsistency in the constraints can be reduced. (4) Artifacts caused by the spurious lobes can be reduced under the condition that the interspacing of transducer elements is larger than the wavelength.

Session 2pUW

Underwater Acoustics: Propagation in Sediments and Inverse Methods

Kyle M. Becker, Chair
Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804-0030

Chair's Introduction—1:00

Contributed Papers

1:05

2pUW1. A two-way coupled mode simulation of sound propagation in a shallow water wedge. T. W. Yudichak, D. P. Knobles, and R. A. Koch (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029)

Near a shoreline an acoustic waveguide can be modeled by a wedge-shaped geometry with a penetrable bottom. The propagation of sound in such a waveguide was studied in a tank experiment conducted by Tindle *et al.* [J. Acoust. Soc. Am. **81**, 275–286 (1987); **81**, 287–294 (1987)]. An analysis of the time series data measured in this experiment was carried out by Royal *et al.* (unpublished), in which time series were simulated using a broadband, one-way, coupled normal mode approach. This analysis is extended using a two-way propagation approach that solves the integral equation form of the coupled mode equation. Also, the modal continuum is treated more realistically through the employment of a leaky mode decomposition. In particular, the importance of backscattering and coupling to the continuum is gauged by comparing the experimental data with the results of the two approaches.

1:20

2pUW2. Extensions to the method of complex images. John A. Fawcett (Defence Res. Establishment Atlantic, Dartmouth, NS B2Y 3Z7, Canada)

In an earlier paper [J. A. Fawcett, J. Acoust. Soc. Am. **108**, 2791–2795 (2000)] the application of the method of complex images to approximating the half-space acousto-elastic Green's function was described. In this paper, extensions to this theory are presented. It is shown that the method can be used for layered half-spaces. For a uniform half-space, analytic time domain expressions are derived for a propagating pulse. By using the complex images in a manner analogous to image theory for a perfect waveguide, the method can be used to approximate a waveguide's Green's function. The Green's function described by the complex image method is accurate for near-field to far-field ranges and is ideal for many propagation and scattering modeling applications.

1:35

2pUW3. T-wave sources, slopes, rough bottoms and continuum. Robert I. Odom and Darin J. Soukup (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, odom@apl.washington.edu)

Bathymetry plays a strong role in the excitation of T-waves by breaking strict mode orthogonality and permitting energy from higher order modes to couple to the lower order modes comprising the T-phase. Observationally (Dziak, 2001) earthquakes with a strong strike-slip component are more efficient at generating T-waves than normal fault mechanisms with the same moment magnitude. It is shown that fault type and orientation correlates strongly with T-wave excitation efficiency. For shallow sources, the discrete modes contribute to the majority of the seismic source field, which is then scattered into the acoustic modes by irregular bathymetry. However, the deeper the earthquake source, the more important the continuum component of the spectrum becomes for the excitation. Deterministic bathymetry and random roughness enter the modal scattering theory as separate terms, and allow the relative contributions from the slope conversion mechanism and bottom roughness to be directly compared. [Work supported by the National Ocean Partnership Program.]

1:50

2pUW4. Electrokinetic transduction of slow waves in porous media. Gareth Block and Nicholas P. Chotiros (Appl. Res. Labs., P.O. Box 8029, Univ. of Texas, Austin, TX 78713-8029, gblock@arlut.utexas.edu)

One candidate for modeling wave propagation in ocean sediments is Biot theory. It predicts two kinds of compressional waves: a fast wave, described by in-phase motion of the pore fluid and porous solid structure, and a slow wave, described by out-of-phase motion. Slow wave motion is heavily damped and has been difficult to measure directly using acoustic transducers. An electrokinetic (EK) transduction technique for studying slow waves is being developed which takes into account the surface chemistry of the grains and pore fluid salinity. EK measurements are advantageous because they are linked directly to out-of-phase, relative motion between the pore fluid and grain structure. EK transmission occurs when an applied voltage drives ionic (and thus fluidic) currents in an electrolyte-saturated material. Conversely, EK reception is possible when fluid electrolyte motion in this material creates a measurable streaming potential. Progress on the experimental apparatus and theoretical model for the EK transmission case will be reported. [Work supported by ONR, Ocean Acoustics.]

2:05

2pUW5. Seismo-acoustic inversion of shallow water data using genetic algorithms. Timothy H. Ruppel (Naval Res. Lab., Stennis Space Center, MS 39529, tim.ruppel@nrlssc.navy.mil)

Peter Gerstoft's seismo-acoustic inversion system uses a Genetic Algorithm to search the parameter space and to estimate the errors in the inversion results, and allows the selection of one of several forward models in the inversion process. The current paper tests this system using data from an experiment conducted off the coast of Panama City by Steve Stanic in the summer of 2001. In the experiment, a vertical array was placed in shallow (6.62 m depth) water and ensonified by a submerged source at distances ranging from 20 m to 90 m. In addition, a number of environmental measurements were taken, making this a suitable data set for acoustic inversion. The SAGA system will be used with several forward models for a comparison. [Work supported by the ONR Program Element No. 62435N.]

2:20

2pUW6. A generalized inversion method: Simultaneous source localization and environmental inversion. Tracianne B. Neilsen and David P. Knobles (Appl. Res. Labs., Univ. of Texas, Austin, P.O. Box 8029, Austin, TX 78713-8029, neilsen@arlut.utexas.edu)

The problem of localizing and tracking a source in the shallow ocean is often complicated by uncertainty in the environmental parameters. Likewise, the estimates of environmental parameters in the shallow ocean obtained by inversion methods can be degraded by incorrect information about the source location. To overcome both these common obstacles—environmental mismatch in matched field processing and incorrect source location in geoacoustic inversions—a generalized inversion scheme is developed that includes both source and environmental parameters as unknowns in the inversion. The new technique called systematic decoupling using rotated coordinates (SDRC) expands the original idea of rotated coordinates [M. D. Collins and L. Fishman, J. Acoust. Soc. Am. **98**, 1637–1644 (1995)] by using multiple sets of coherent broadband rotated coordinates, each corresponding to a different set of bounds, to systematically decouple the unknowns in a series of simulated annealing inversions. The results of applying the SDRC inversion method to data from the Area Characterization Test II experiment performed on the New Jersey continental shelf are presented. [Work supported by ONR.]

2:35

2pUW7. Evaluation of an autoregressive spectral estimation technique for determining horizontal wave-number content in shallow water. Kyle M. Becker[a] (MIT/WHOI Joint Prog. in Oceanogr. and Oceanogr. Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543) and George V. Frisk (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Modal-based techniques for geoacoustic inversion require estimates of discrete horizontal wave numbers corresponding to the propagating modes in a shallow-water waveguide. For range-dependent waveguides, local wave-number content can be extracted as a function of range using methods analogous to short-time Fourier transform techniques as applied to non-stationary time series data. A limiting factor in these techniques is the aperture length, L, required to resolve closely spaced horizontal wave-number content. Using classical spectral estimation methods, the limit of resolution is proportional to $1/L$. In order to increase the spatial resolution of local horizontal wave-number estimates, the use of a high-resolution autoregressive spectral estimator was investigated for determining horizontal wave-number content. The performance of the estimator was evaluated numerically using both time-series and spatial data. Results are presented of estimator performance for various aperture lengths using signals with increasing levels of additive noise and signals with closely spaced wave-number components. Comparisons are made with results of classical methods as well as other high-resolution methods. The results show the AR method to be well suited for wave-number estimation in shallow-water waveguides. [Work supported by ONR.] [a] Presently with the Acoustics Group, The Pennsylvania State Univ., Appl. Res. Lab., P.O. Box 30, State College, PA 16804, kmb166@psu.edu.

2:50

2pUW8. Sediment tomography in the East China Sea: Compressional wave speed and attenuation inversions from mode travel time dispersion and Airy phase measurements.[a] Colin J. Lazauski, Gopu R. Potty, James H. Miller, Chuen-Song Chen (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882, lazauski@oce.uri.edu), and Peter H. Dahl (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

This paper discusses ongoing data analysis results from the acoustic bottom interaction experiment conducted in May–June 2001 in the East China Sea as part of the Asian Seas International Acoustics Experiment (ASIAEX-2001). Using time-frequency scalograms of broadband signals, the modal arrivals and group speed minimums (Airy Phase) of several modes are clearly observed. The structure of the Airy Phase signal forms the basis of this inversion technique. Utilizing the Airy Phase group speed minimums and corresponding pressure amplitudes of each observable mode, the sediment compressional wave speed and attenuation as a function of depth are derived. The group speed minimum for each mode provides additional information on the compressional wave speed in the modal sediment depth penetration interval. Inverted speeds and estimated modal penetration depths are then used to develop the sediment profile. The estimated resolution is dependent on the number and frequency span of the observable modes. This method has the advantage of utilizing only one hydrophone to obtain rapid estimates of sediment properties. Additionally, it also complements other inversion schemes that can benefit from accurate a priori environmental information by limiting the size of the search parameter space. Estimated sediment properties from several areas are presented with verification from coring results. [Work supported by ONR.] [a] For Acoustical Oceanography Best Student Paper Award.

Session 3aAAa

Architectural Acoustics: Computer Modeling in Acoustical Consulting II

Lily M. Wang, Chair
Architectural Engineering, University of Nebraska, 200B Peter Kiewit Institute, 1110 South 67th Street, Omaha, Nebraska 68182-0681

Invited Papers

8:00

3aAAa1. Validation of an auralization system. Claus Lynge Christensen, Christoffer A. Weitze, Jens Holger Rindel, and Anders Christian Gade (Ørsted·DTU, Acoust. Technol., Tech. Univ. of Denmark, Bldg. 352, Ørsted Plads, 2800 Kgs. Lyngby, Denmark, clc@oersted.dtu.dk)

The room acoustics program ODEON provides auralization using fully filtered binaural room impulse responses, each reflection being filtered through nine octave bands and a set of head-related transfer functions. Using the full filtering scheme allows, in principle, a complete audible presentation of all the properties, time-variant frequency coloration, as well as directional information predicted by the room acoustics program. Two methods of verification have been applied. The first method is based on direct measurements on the impulse responses predicted by ODEON, using the room acoustics measuring system DIRAC in order to verify that the auralization method is actually capable of reproducing the predicted room acoustic parameters. Monaural auralization filters were used for this purpose. The other method is an audible comparison between *in situ* recordings of a singing person in real rooms and the ODEON auralization of the same situations. The latter verification is part of an ongoing European research project, CAHRISMA, on restoration of the acoustics in old Byzantine churches and mosques in Istanbul.

8:30

3aAAa2. Effect of model detail level on room acoustic computer simulations. David T. Bradley and Lily M. Wang (Univ. of Nebraska-Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, dbradley@mail.unomaha.edu)

Computer modeling of room acoustics is limited due to approximations regarding wavelength to surface size ratios and diffraction. In fact, it is suggested that a model with high geometrical fidelity may actually lead to less accurate predictions. This project seeks to clarify how the level of model detail affects the accuracy of these simulations. Three models of the University of Nebraska's Strauss Recital Hall have been generated at varying levels of detail; low (a rough model of the essential architectural dimensions), medium (a more accurate representation of the space), and high (a complete rendering of the surfaces within the space). The room acoustics modeling program Odeon is used to calculate objective measures (i.e., reverberation time and clarity) and create auralizations for each model. Meanwhile, the actual acoustical properties of Strauss Hall have been measured. The objective measure results are compared between the models and against measured data. Additionally, subjective tests are run with auralizations from all three models and the original space to determine if the difference in level detail can be perceived psychoacoustically. Quantitative results from these comparisons are presented to help users of room acoustic simulation software understand how the level of model detail will produce maximum accuracy.

9:00

3aAAa3. The development and use of the diffusion scattering coefficients in room modeling software: The effect of diffuse reflections on reverberation time. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774) and Trevor J. Cox (Univ. of Salford, Salford M5 4WT, UK)

In the last 20 years, much progress has been made in the measurement and characterization of sound diffusing surfaces. This presentation describes how this research might improve the accuracy of computer based geometric room modeling. To model diffuse reflections, one needs to know the scattering distribution and the portion of scattered energy that is diffusely reflected. AES Working Group SC-04-02 recently published an information document AES-4id-2001, which describes a measurement procedure to determine the polar balloons from scattering surfaces. This allows programmers to use actual measured or calculated distributions instead of Lamberts law, which applies to incoherent optical diffusion. ISO WG 25 is currently addressing a measurement procedure to determine the random incidence scattering coefficient. This scattering coefficient is used directly in geometrical modeling programs. The two measurement procedures will be contrasted. It will then be demonstrated by example how the use of diffusion data can affect the predicted reverberation time.

9:30

3aAAa4. Uses and limitations of statistical and geometrical modeling of nonexponential decay curves in performance spaces. Jason E. Summers, Rendell R. Torres, Yasushi Shimizu (Prog. in Architectural Acoust., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180), and J. Christopher Jaffe (Jaffe Holden Acoust., 114A Washington St., Norwalk, CT 06854)

The conditions under which the high-frequency, diffuse-field model of coupled rooms is a valid approximation to geometrical acoustics have been examined by comparison with ray-tracing predictions of decay curves in two and three room systems. Results particular to coupled rooms were noted. Because nonexponential decay shape is sensitive to both decay rates and relative energy densities, corrections to these parameters meaningfully improved the diffuse-field model. Variations in the free-path distribution of each subroom, introduced by coupling, complicated the use of improved decay models. Also, the expected decrease in energy density with distance from the source was found to result in spatial dependence of decay shape for certain coupling geometries. Insights from this study were used in the construction of diffuse-field and geometrical computer models of Bass Hall, a 2056 seat multipurpose auditorium with an acoustically coupled stage house. Preliminary results indicate that high-frequency decay curves in each of the subrooms predicted by geometrical acoustics are well matched to the predictions of diffuse-field models. The use of both models as prediction and design tools is assessed by comparison at high frequencies with measurements made during occupied and unoccupied conditions. [Research supported by the Bass Foundation.]

10:00

3aAAa5. Case studies of computer model applications in consulting practice. Gary Siebein, Hyun Paek, Mark LoRang, and Courtney McGuinnes (Siebein Assoc., Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607)

Six case studies of computer model applications in a consulting practice will be presented to present the range of issues that can be studied with computer models as well as to understand the limitations of the technique at the present time. Case studies of elliptical conference rooms demonstrate basic acoustic ray principles and suggest remediation strategies. Models of a large themed entertainment venue with multiple amplified sound sources show how visualization of the acoustic ray paths can assist a consultant and client in value engineering locations and amounts of acoustic materials. The acoustic problems with an angled ceiling and large rear wall were studied when an historic church was converted to a music performance hall. The computer model of an historic hall did not present enough detailed information and was supplemented with physical model studies and full size mock-up tests of the insertion of an elevator door that would open directly into the concert room. Studies to demonstrate the amount of room model detail to obtain realistic auralizations were also conducted. The integration of architectural acoustic design and audio system design were studied in computer models of a large church sanctuary.

WEDNESDAY MORNING, 5 JUNE 2002 KINGS GARDEN SOUTH, 10:30 A.M. TO 1:00 P.M.

Session 3aAAb

Architectural Acoustics: Student Design Competition

Lily M. Wang, Cochair
Architectural Engineering, University of Nebraska, 200B Peter Kiewit Institute, 1110 South 67th Street, Omaha, Nebraska 68182-0681

Robert C. Coffeen, Cochair
School of Architecture and Urban Design, University of Kansas, Marvin Hall, Lawrence, Kansas 66045

The Technical Committee on Architectural Acoustics of the Acoustical Society of America and the National Council of Acoustical Consultants are sponsoring a student design competition which will be professionally judged at the meeting.

The purpose of this design competition is to encourage students enrolled in Architecture, Architectural Engineering, and other University curriculums that involve building design and/or acoustics to express their knowledge of architectural acoustics and building noise control in the schematic design of portions of a building where acoustical considerations are of primary importance.

The submitted designs, which will be displayed at the Pittsburgh ASA Meeting, will be judged by a panel of professional architects and acoustical consultants. An award of $1,000 will be made to the submitter(s) of the design judged ''first honors.'' Four awards of $500 each will be made to the submitters of four entries judged ''commendation.''

Session 3aABa

Animal Bioacoustics: Acoustics of Terrestrial Mammals and Birds

Ann E. Bowles, Chair
Hubbs-Sea World Research Institute, 2595 Ingraham Street, San Diego, California 92109

Chair's Introduction—8:00

Invited Papers

8:05

3aABa1. General principles involved in the effect of noise on hearing and vocal communication in birds. Robert J. Dooling and Michael L. Dent (Dept. of Psych., Univ. of Maryland, College Park, MD 20742, dooling@psyc.umd.edu)

Birds provide very useful models for understanding the effects of noise on hearing and acoustic communication. They are excellent subjects for laboratory studies of hearing in which signals and noise can be precisely defined and delivered and behavioral responses can be unambiguously interpreted. For this reason, a huge amount is already known about their hearing. Acoustic communication is critically important for most species of birds and some even acquire their communication signals through vocal learning. For this reason, a lot is already known about how birds perceive complex acoustic signals such as vocalizations. Drawing from both field and laboratory studies, we review what is known about the effects of noise on hearing and vocal communication in birds. This includes the effects of intense noise on the ear, the effects of background noise on the detection and discrimination of both simple sounds and complex vocalizations, and the spatial effects of signal detection in noise in the free-field. As a whole, these studies show that birds are resistant to damage and interference from noise and have developed a variety of strategies to effectively communicate.

8:25

3aABa2. Analysis of auditory spatial receptive fields: An application of virtual auditory space technology. Terry T. Takahashi, Clifford H. Keller, David R. Euston, and Michael L. Spezio (Inst. of Neurosci., Univ. of Oregon, Eugene, OR)

Virtual auditory space technology, typically used to simulate acoustical environments, also allows one to vary one sound localization cue independently of others. VAST was used to determine the contributions of interaural time and level differences (ITD, ILD) to the spatial receptive fields (RFs) of neurons in the owl's midbrain. The presentation of noise filtered so that only ITD varied evoked a response along a vertical strip of virtual space, called the ITD-alone RF. Conversely, when ITD was fixed at the cell's optimum and the ILD spectrum of each location was presented, the cell responded along a horizontal strip, called the ILD-alone RF. The spatial RF was at the intersection of the ITD and ILD-alone RFs. The cell's ILD tuning across frequency, combined with individualized head-related transfer functions, was transformed into an ILD-alone RF that predicted half the variance in the measured one. This discrepancy was due partly to the poor response of the neurons to tones, and a new method of inferring frequency-specific ILD tuning from responses to noise explained about 75% of the variance. By understanding how spatial RFs are constructed, it is possible to infer the neural image of complex auditory scenes containing multiple sources and echoes. [Work supported by NIDCD.]

8:45

3aABa3. Sex differences in razorbill (Family: *Alcidae*) parent–offspring vocal recognition. Stephen J. Insley (Hubbs-Seaworld Res. Inst., 2595 Ingraham St., San Diego, CA 92109), Rosana Paredes Vela, and Ian L. Jones (Memorial Univ. of New Foundland, Canada)

In this study we examines how a pattern of parental care may result in a sex bias in vocal recognition. In Razorbills (*Alca torda*), both sexes provide parental care to their chicks while at the nest, after which the male is the sole caregiver for an additional period at sea. Selection pressure acting on recognition behavior is expected to be strongest during the time when males and chicks are together at sea, and as a result, parent–offspring recognition was predicted to be better developed in the male parent, that is, show a paternal bias. In order to test this hypothesis, vocal playback experiments were conducted on breeding Razorbills at the Gannet Islands, Labrador, 2001. The data provide clear evidence of mutual vocal recognition between the male parent and chick but not between the female parent and chick, supporting the hypothesis that parent–offspring recognition is male biased in this species. In addition to acoustic recognition, such a bias could have important social implications for a variety of behavioral and basic life history traits such as cooperation and sex-biased dispersal.

9:05

3aABa4. Identifying bird species from bird song using frequency component analysis. Jacky Mallett and Irene Pepperberg (Media Lab., MIT, 20 Ames St., Cambridge, MA 02139)

The species-distinctive and highly stereotyped nature of bird vocalizations makes birdsong an obvious candidate for computational analysis. By using techniques from previous work in automatic music recognition and auditory analysis, a database of the detailed auditory characteristics of a selection of songs from over 150 different birds from the North Eastern region of the United States of America has been created. Examination of this database has allowed a picture to be built up of species-distinctive characteristics, and

provides a more general picture of bird vocalizations than has previously been available. From this work it appears that frequency use by birds is of some significance, with evidence that the presence of a particular set of dominant frequencies in a bird's song can be used as a distinctive fingerprint for species recognition. This paper reviews previous work in this area, outlines general computational methods used to extract this information, and presents results of automatic bird recognition using characteristics derived from the database.

9:25

3aABa5. Acoustic surveys for Mexican spotted owls (*Strix occidentalis lucida*): An analysis. Ann E. Bowles, Chris Martindell (Hubbs-Sea World Res. Inst., 2595 Ingraham St., San Diego, CA 92109, annb1@san.rr.com), Kenneth J. Plotkin, Bruce Ikelheimer (Wyle Labs., Ste. 701, 2001 Jefferson Davis Hwy., Arlington, VA 22202), and Tim Lavallee (GeoMarine, Inc., 11846 Rock Landing Dr., Newport News, VA 23606)

During acoustic surveys for Mexican spotted owls, the effective detection range is presumed to be 0.25 mi (433 m). However, variations within and between surveys are observed, leading to a variance in owl density estimates. While owl behavior may explain some variation, topography and ambient noise are also likely to be important. To determine the influence of these factors, data from acoustic surveys for Mexican spotted owls in the Gila National Forest (April–July, 2000–2001) were examined. Measurements of owl and human call levels were made with a Sony TCD-10 Pro II DAT equipped with an ACO 7013 microphone. Ambient noise was collected using 40 Larson-Davis 820 and 824 sound level meters in owl activity centers. Wyle Laboratories NMSIM software was used to model propagation of owl and human calls. Owls produced calls with estimated maximum source levels of 92–98-dB SPL. Human callers produced maximum source levels of 88–95-dB SPL. Detection was possible out to more than 2 km under ideal conditions, but topography and ambient noise had a large effect. Corrections for these factors would greatly improve estimates of area surveyed, and thus owl density estimates. [Work supported by the U.S. Air Force, Air Combat Command.]

9:45–9:55 Break

Contributed Papers

9:55

3aABa6. Frequency selectivity in canaries with a hereditary hearing loss. Amanda M. Lauer and Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742, alauer@psyc.umd.edu)

Sensorineural hearing loss is associated with reduced frequency selectivity due to the broadening of the auditory filters in mammals. In European starlings, hearing loss caused by the ototoxic drug kanamycin results in the widening of auditory filters at 5 kHz [Marean *et al.* (1998)]. In the present study, we examine frequency selectivity in a bird with a permanent hereditary hearing impairment, the Belgian Waterslager (BWS) canary. This strain of canary has long been bred for its loud, low-frequency song, and has been shown to have a hearing loss primarily at higher frequencies (2 kHz and above). Using operant conditioning and the method of constant stimuli, thresholds for detecting pure tones in flat-spectrum broadband noise were measured in BWS and non-BWS canaries. Critical ratios were calculated for comparison with other species of birds. At higher frequencies, critical ratios for BWS canaries were much larger than those of non-BWS canaries and other birds, suggesting reduced frequency selectivity in the region of the birds' hearing loss. [Work supported by NIDCD R01DC001372 to RJD and Brenda M. Ryals.]

10:10

3aABa7. Narrowband masking in birds. Stacey L. Brown, Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742, sbrown@psyc.umd.edu), and Marjorie R. Leek (Walter Reed Army Medical Ctr., Washington, DC 20307)

Hearing studies show that zebra finches, *Taeniopygia guttata*, small song birds that produce broadband, harmonic vocalizations are unusually sensitive to the mistuning of a single component in a harmonic complex [Lohr and Dooling, J. Comput. Psychol. **112**, 36–47 (1998)]. Such a change in a harmonic stimulus creates both spectral and temporal cues. Previously it has been reported that these birds have very good temporal resolution of fine detail within harmonic complexes [Dooling *et al.*, Hear. Res. **152**, 159–172 (2001)]. Here the spectral resolving power is examined by measuring auditory patterns of masking by narrowband noises. Masking patterns were measured in zebra finches using operant conditioning and the method of constant stimuli for a 100-Hz band noise masker centered at the birds frequency of best hearing, 2860 Hz. Probe frequencies were tones between 1000 Hz and 5700 Hz. Results demonstrate a symmetrical masking pattern with less upward spread of masking and a narrower resolution bandwidth in zebra finches than has been observed in humans. These findings are consistent with other measures of avian spectral resolution. [Work supported by NIH Grant DC-00198 to RJD.]

10:25

3aABa8. Computer synthesis of bird songs and calls. Mark Kahrs (Dept. of Elec. Eng., Univ. of Pittsburgh, 348 Benedum Hall, Pittsburgh, PA 15261) and Federico Avanzini (Universitá degli Studi di Padova, 35131 Padova, Italy)

We describe the computer simulation of a one-mass source together with a simple transmission line model for a psittacine bird. The syrinx is modeled as a lumped mass subject to elastic restoring forces and internal dissipation. Nonlinear interaction with the airflow follows the Ishizaka and Flanagan glottal model. This is also used to model complete closure of the syrinx: during closure an additional restoring force is added and dissipation is increased. During the whole closed phase the syringeal flow is zero, and consequently its spectrum is broadened and higher partials are generated. The vocal tract transmission lines are implemented with fractional delay lines and the transmission lines are assumed lossless. The beak is implemented as two complementary fifth order Butterworth filters. The pole trajectory for the filters approximates the nonlinear path of the beak cutoff as discussed by Fletcher. [M.K. was supported by the Fulbright Foundation as well as Tekes. F.A. was funded by the Sound Source Modeling Project. Both authors were supported in part by the Academy of Finland.]

10:40

3aABa9. Discrimination of dynamic moving ripples in the zebra finch, *Taeniopygia guttata*. Michael S. Osmanski, Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742), and Didier A. Depireux (Univ. of Maryland at Baltimore, Baltimore, MD)

Auditory processing of complex broadband sounds known as moving ripples has been studied both physiologically at the cortical level in mammals and psychophysically in humans [Depireux *et al.* (2001); Chi *et al.* (1999)]. These stimuli share spectro-temporal properties with many natural sounds, including species-specific vocalizations and the formant transitions of human speech [Versnel and Shamma (1998)]. One test of the generality of ripple processing beyond mammals would be to examine a non-mammalian species. Zebra finches may be excellent subjects for such

a study because they produce complex broadband harmonic songs and neuronal responses in their auditory forebrain may be exquisitely tuned to the specific spectro-temporal patterns of their songs [Theunnissen and Doupe (1998)]. We trained these birds to discriminate between flat-spectrum broadband noise and moving ripples of different densities that move up or down in frequency at various rates. Results show that discrimination in zebra finches is better at those ripple densities and velocities which are prominent in their species-specific harmonic vocalizations. [Work supported by NIH Grant No. DC-00198 to RJD and NIDCD Training Grant No. DC-00046.]

Session 3aABb

Animal Bioacoustics: General Topics in Animal Bioacoustics

Ann E. Bowles, Chair
Hubbs-Sea World Research Institute, 2595 Ingraham Street, San Diego, California 92109

Chair's Introduction—10:55

Contributed Papers

11:00

3aABb1. Acoustical and functional analysis of Mountain lion (*Puma concolor*) vocalizations. Jacquelyn Potter (Dept. of Biological Sci., Western Illinois Univ., 1 University Circle, Macomb, IL 61455, jacquelyn_potter@yahoo.com)

A 2-year study resulted in acoustic analysis of the structure of over 900 mountain lion vocalizations recorded in a seminatural setting at Wildlife Prairie Park near Peoria, Illinois. A vocal repertoire was obtained by describing quantitative variables about the sounds, i.e., frequency of the dominant part of the sound (beginning, ending, maximum, and minimum), duration, and number of components. Other variables described the tonal, harmonic, and wideband qualities of the sounds. Behavioral data were collected during the same period. Further analysis of both acoustic and behavioral data was completed to develop a correlation matrix between vocalizations and behavior. This study also looked at the effects of seasons on vocal behavior. Correlations were found between vocalization types and rates of usage with specific behaviors. Vocalization type and the usage rate also varied by season.

11:15

3aABb2. Acoustic correlates of human responses to domestic cat (*feliscatus*) vocalizations. Nicholas Nicastro (Dept. of Psych., Cornell Univ., 243 Uris Hall, Ithaca, NY 14853)

As part of ongoing research on coevolution of vocal communication between humans and domestic cats, perceptual data were collected on participants as they listened to recorded cat vocalizations. In experiment 1, human subjects were asked to rate the pleasantness of 100 meows along a 7-point scale, from most to least pleasant. In experiment 2, a different group of participants was asked to rate the urgency of the same meows along a 7-point scale, from most to least urgent. Linear regression analysis of the results showed a strong inverse correlation between pleasantness and urgency. Acoustic correlates of pleasantness included reduced frequency modulation, a downward shift in fundamental frequency, and fewer noisy segments. Correlates of urgency included increased duration, higher Wiener entropy, and acute spectral tilt. It is speculated that humans' affective responses to these acoustic qualities, in conjunction with contextual cues, may form the basis of the communication of more specific meanings by cats to humans.

11:30

3aABb3. Extraction of directional pitch change by Japanese monkeys. Tokuro Takahashi and Hiroshi Riquimaroux (Dept. of Knowledge Eng. & Computer Sci., Doshisha Univ., Kyotanabe, Kyoto 610-0321, Japan)

Perception of directional pitch change was behaviorally investigated in Japanese monkeys on an operant Go–Nogo task with water reward. Results show that the difference in the fundamental frequency makes a difference in direction of pitch change. A pair of tone bursts and/or harmonically structured complex tone bursts were sequentially presented. The monkeys could always discriminate the direction of pitch change by using the fundamental frequency cue but in a specific case. Their performance was worse than a chance level, or they oppositely responded to the fundamental frequency cue in the specific case. The case was that the temporal sequence was made of a simple tone burst and a complex tone burst where the frequency of the simple tone burst was in between the fundamental and the second harmonic frequencies of the complex tone burst, and the fundamentals of the complex tone were lower than 500 Hz. They showed amelioration in using the fundamental frequency cue as interval between two tones was expanded. The results were confirmed with probe tests. [Research supported by Special Coordination Funds, and a grant to RCAST at Doshisha University from the Ministry of Education, Culture, Sports, Science and Technology of Japan.]

11:45

3aABb4. Sound duration as a perceptual cue influencing vocal behavior of male bullfrogs. Andrea M. Simmons (Dept. of Psych., Brown Univ., Providence, RI 02912, Andrea_Simmons@brown.edu)

Female frogs of several species use the temporal cue of sound duration to aid in mate choice. Little is known, however, about the sensitivity of male frogs to this cue. Male bullfrogs emit a complex advertisement call to attract females for mating, and to announce territory occupation to other males. In two experiments, the sensitivity of vocalizing male bullfrogs to field playbacks of advertisement calls differing in duration was examined. The number and latency of evoked vocal responses to the stimuli was used as a measure of perception. Males responded with fewer calls, at longer latencies, to stimuli shorter in duration than the standard signal (with a duration at the mean value for the species). Males preferred stimuli longer in duration than the standard signal, responding with more calls at shorter latencies. They did not, however, significantly lengthen their own calls in response to playbacks of long duration signals. This preference for ''supernormal'' stimuli may be an important factor mediating the evolution of communication signals. [Work supported by NIH.]

12:00

3aABb5. Acoustic communication by ants. Robert Hickling (Sonometrics, Inc., 8306 Huntington Rd., Huntington Woods, MI 48070)

Many ant species communicate acoustically by stridulating, i.e., running a scraper over a washboard-like set of ridges. Ants appear to be insensitive to airborne sound. Consequently, myrmecologists have concluded that the stridulatory signals are transmitted through the substrate. This has tended to diminish the importance of acoustic communication, and it is currently believed that ant communication is based almost exclusively on pheromones, with acoustic communication assigned an almost nonexistent role. However, it can be shown that acoustic communication between ants is effective only if the medium is air and not the substrate. How, then, is it possible for ants to appear deaf to airborne sound and yet communicate through the air? An explanation is provided in a paper [R. Hickling and R. L. Brown, ''Analysis of acoustic communication by ants,'' J. Acoust. Soc. Am. **108**, 1920–1929 (2000)]. Ants are small relative to the wavelengths they generate. Hence, they create a near field, which is characterized by a major increase in sound *velocity* (particle velocity of sound) in the vicinity of the source. Hair sensilla on the ants' antennae respond to sound velocity. Thus, ants are able to detect near-field sound from other ants and to exclude extraneous airborne sound.

WEDNESDAY MORNING, 5 JUNE 2002 RIVERS ROOM, 8:45 TO 11:00 A.M.

Session 3aMU

Musical Acoustics: General Topics in Musical Acoustics

Uwe J. Hansen, Chair
Department of Physics, Indiana State University, Terre Haute, Indiana 47809

Contributed Papers

8:45

3aMU1. The effect of inharmonic partials on pitch of piano tones. Brian E. Anderson (Dept. of Phys., Brigham Young Univ., Provo, UT 84602, loudspeakerdesign@hotmail.com)

Piano tones have partials whose frequencies are sharp relative to harmonic values. A listening test was conducted to determine the effect of inharmonicity on pitch for piano tones in the lowest three octaves of a piano. Nine real tones from the lowest three octaves of a piano were analyzed to obtain frequencies and amplitude of their partials. Synthetic inharmonic tones were produced from these results. Synthetic harmonic tones, each with a twelfth of a semitone increase in the fundamental, were also produced. A jury of 10 listeners matched each synthetic inharmonic tone to one of the synthetic harmonic tones. The effect of the inharmonicity on pitch was determined from an average of the listeners' results. For the nine piano tones studied, inharmonicity increased pitch from a twelfth of a semitone to nearly a semitone.

9:00

3aMU2. Prosodic persistence in music performance and speech production. Melissa K. Jungers, Caroline Palmer (Dept. of Psych., Ohio State Univ., 1885 Neil Ave., Columbus, OH 43210, jungers.2@osu.edu), and Shari R. Speer (Ohio State Univ., Columbus, OH 43210)

Does the rate of melodies that listeners hear affect the rate of their performed melodies? Skilled adult pianists performed two short melodies as a measure of their preferred performance rate. Next they heard, on each trial, a computer-generated performance of a prime melody at a slow or fast rate (600 or 300 ms per quarter-note beat). Following each prime melody, the pianists performed a target melody from notation. The prime and target melodies were matched for meter and length. The rate of pianists' target melody performances was slower for performances that followed a slow prime than a fast prime, indicating that pianists' performances were influenced by the rate of the prime melody. Performance duration was predicted by a model that includes prime and preferred durations. Findings from an analogous speech production experiment show that a similar model predicts speakers' sentence rate from preferred and prime sentence rates. [Work supported by NIMH Grant 45764 and the Center for Cognitive Science.]

9:15

3aMU3. Sound of a cup with and without instant coffee. Andrew Morrison and Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115)

An empty coffee cup, like an ancient Chinese two-tone bell, emits two distinctly different tones, depending upon where it is tapped. When it is filled with hot water, and some instant coffee is added, however, a whole new set of sounds is heard when the cup is tapped. The pitch rises an octave or more as the foam clears due to the dramatic change in the speed of sound in the bubble-filled liquid. A similar, but smaller, effect was noted in beer by Bragg [*The World of Sound* (1968)] and in hot chocolate by Crawford [Am. J. Phys. (1982)]. We describe the modes of vibration in a coffee cup and the sound emitted by a coffee cup as filled with instant coffee as the bubble density changes.

9:30

3aMU4. Computer-animated illustrations of vibrations and waves. W. Bruce Richards (Oberlin College, Oberlin, OH 44074)

Undergraduate students in a general-education course on musical acoustics often have difficulty understanding the actual time-domain motions of vibrating systems which textbooks can only depict using static diagrams. In particular, students have difficulty relating the longitudinal average displacements of gas molecules in acoustic waves to (transverse) graphs of these displacements as a function of position. To better convey these ideas, a number of animated graphic diagrams have been programmed on a Macintosh computer. For longitudinal waves, these include

one-dimensional standing waves in cylindrical pipes with open or closed ends, various pulse waves reflecting from the open or closed end of a cylindrical pipe, and standing waves in a conical pipe with an open end. In addition, the transverse motions of bowed, plucked, and hammered strings have been programmed. Presented in every case is an exaggerated, slow motion picture of the system's displacement as it evolves in time. If desired, an adjacent coordinated and animated graph of displacement, pressure, or other dynamical variable may also be displayed. As part of this paper the computer programs will be demonstrated.

9:45

3aMU5. Drumhead contact time measurement using metallic leaf. José Sánchez and James Irwin, Jr. (Dept. of Elec. & Computer Eng., 1501 W. Bradley Ave., Bradley Univ., Peoria, IL 61625, jhirwin@bradley.edu)

This work investigates the relationship between mass, diameter, impact location, and energy of steel balls impacting upon the surface of the drumhead, and the time they remain in contact. The contact time is measured by making the ball and metallic leaf contacts in an electrical switch. The leaf was applied to the drumhead surface and connected to the circuit through a wire. The balls were connected to the circuit using a solder-attached hair-fine wire. Contact time was measured at a constant drumhead tension at three distinct locations on the drumhead: the center, halfway between the center and the rim, and an inch from the rim. In addition, a single series was measured from the center to the rim at $\frac{1}{2}$-in. intervals. Initial work indicates that the contact time is not dependent upon energy, impact location, or diameter, but only upon the mass of the steel ball. The relationship of the mass-dependent contact time to energy transfer and to impact spectra is developed, showing mass is the determining parameter needed for the initial conditions used to simulate drumstick/drumhead interaction (impact effects).

10:00

3aMU6. Mode studies in triangular bellplates. Uwe J. Hansen (Dept. of Phys., Indiana State Univ., Terre Haute, IN 47803), Daniel Lavan, and Suzanne Hogg (Tech. Univ., Sydney, Australia)

Preliminary results from mode studies in triangular aluminum bellplates using electronic holographic interferometry were reported at the 136th ASA meeting [Uwe J. Hansen, J. Acoust. Soc. Am. **104**, 1767(A) (1998)]. Since then the TU Sydney co-authors of this paper have performed finite element calculations with results which are remarkably congruent with the holographic data. The first 13 modes will be discussed and shown in holographic as well as FEM representation.

10:15

3aMU7. Shape of a clamped stiff harpsichord wire driven at a resonant frequency. Roger J. Hanson, Hilliard Kent Macomber, and Mathew A. Boucher (Dept. of Phys., Univ. of Northern Iowa, Cedar Falls, IA 50614, roger.hanson@cfu.net)

A wire transversely driven by a sinusoidal force at the resonant frequency of a vibrational mode vibrates at the driving frequency and at harmonics generated by nonlinear processes in the wire. If the amplitude of a harmonic is measured as a function of position along the wire, its shape is revealed. It differs significantly from a sinusoid in the vicinity of either end of the wire because the ends are clamped and the wire has significant stiffness. The shapes of various harmonics have been determined for a brass harpsichord wire, 70 cm long, from optical detector measurements made at different distances from a clamped end. Knowledge of shape facilitates the determination of antinode amplitudes of harmonics when the gross motion of the wire is so large that the detectors must be positioned near an end of the wire because of their very limited dynamic range. Some observations of harmonics and related phenomena were reported previously [Hanson *et al.,* J. Acoust Soc. Am. **108**, 2592 (2000); **106**, 2141 (1999)]. The shape information is also needed to help separate nonlinear effects possibly occurring in the detectors from those of interest, occurring in the wire itself.

10:30

3aMU8. Regimes of oscillation and reed vibrations in lingual organ pipes. Eric Cox and Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115)

We compare the sound generation in six lingual organ pipes including two trompette pipes, two oboes pipes, a krummhorn, and a rankett, the latter two of which were described in a previous paper [Cox and Rossing, paper 2aMU8, 142nd ASA meeting]. With the resonator attached, each pipe sounds in distinct regimes of oscillation, separated by discontinuities. The nominal frequencies of the pipes are generally slightly lower than the resonance frequency of the resonator but slightly higher than the natural frequency of the reed. The air mass loading of both the resonator and shallot lowers the reed frequency. Changing the blowing pressure has little effect on the sounding frequency but a modest effect on sound level.

10:45

3aMU9. Acoustic analysis of an Olmecan whistle. Sergio Beristain (President IMA, Lab. Acoust., Faculty of Elec. and Mech. Eng., Natl. Polytechnic Inst., P.O. Box 75805, 07300 Mexico City, Mexico), Rolando Menchaca (ESCOM, IPN), and Roberto Velazquez (CIC, IPN)

Thousands of stone artifacts over 2500 years of age have been found in the Olmecan area in the southeast region of Mexico. These range from the famous big heads with helmets (about 2 m in diameter), to small pieces with precisely drilled holes, which some archaeologists consider to have some simple uses, much simpler than the work needed to produce the stone artifact itself. The one studied here (about 3 cm in size), is considered by the acoustics community as an air-phone, and a detailed anaysis has been done employing FFT techniques in order to find out the frequency spread, the particular tones produced and the sound radiated power through the different holes and cavities. The artifact is made of a very solid stone, so-called ilmenite, believed to have titanium, which is very hard to drill. Nevertheless, many similar pieces have been found in the area, which means they were built on purpose, and the material used gives the idea of a sacred application. Attempts have been made to reproduce the artifacts, which produce sounds similar to those from the original pieces.

Session 3aNS

Noise, Speech Communication and Animal Bioacoustics: Public Policy for Noise

Lawrence S. Finegold, Chair
Finegold & So, Consultants, 1167 Bournemouth Court, Centerville, Ohio 45459

Invited Papers

8:00

3aNS1. Public policy and the ASA. E. J. Walsh (Boys Town Natl. Res. Hospital, Omaha, NE 68131)

At the Ft. Lauderdale meeting of the Acoustical Society of America, the President of the Society authorized the formation of an ad hoc public policy committee to be named the Panel on Public Policy (PoPP). The Panel's mission is to serve the public interest by addressing and commenting on societal issues that fall within the broad range of disciplines that define the scope of the Society. The Panel's goal is to guide the development of ASA policies on, and responses to, questions that arise in relation to those issues. The operational plan calls for the Panel to execute its responsibility by generating and releasing Executive Council mandated and approved position statements, reports, or results of in-depth studies on issues affecting public health and welfare for use by appropriate legislative and policy-making entities. The structure, composition, and operational plan of the Panel will be discussed in general terms and specific prospective projects will be summarized.

8:10

3aNS2. Rationale for noise regulation: How we got here and why we may not stay. Sanford Fidell (Fidell Assoc., Inc., 23139 Erwin St., Woodland Hills, CA 91367)

Today's framework of environmental noise regulation in the United States is the product of a half century's evolution in understandings of noise effects, and in societal interest and policy interpretations of them. Although it is tempting to associate the evolution of noise policy with rational adaptations to developments in scientific understanding, this is a myopic and ultimately untenable view. It is hardly surprising that little of practical utility for regulatory purposes has come from decades of research in this area, considering the unsystematic and atheoretical nature of much of what has been learned about individual and community response to noise. Today's annoyance-based rationale for regulatory policy is only one of several historical perspectives, and not necessarily the most useful of them. Further changes in the rationale for assessing noise impacts may be expected as court rulings increasingly challenge reliance upon long-term exposure metrics as predictors of noise effects; as the magnitude of non-acoustic contributions to annoyance becomes clearer; and as the information that the U.S. Federal Interagency Committee on Noise last reviewed a decade ago becomes more and more stale.

8:40

3aNS3. National policy on the effects of underwater noise on marine mammals and turtles. Roger L. Gentry (NOAA, Natl. Marine Fisheries Service, Silver Spring, MD 20910)

Marine anthropogenic noise is increasing at an unknown rate yet its effects on mammals and turtles are poorly documented. Such information is central to noise standards for marine animals required by two federal laws (MMPA and ESA). Good policy on underwater noise would set standards that are neither too liberal (thereby threatening the well being of marine animals), nor too conservative (thereby unnecessarily impairing essential human activities at sea). This balance must be found using existing data; standards cannot await new research. Policy must also lay out rules governing the introduction of novel underwater sound sources whose effects on animals are unknown. Should policy forbid such introductions until all possible consequences are known, or does a safe, incremental method of introduction exist? Should exceptions be made for novel sources associated with national defense? Policy must cope with sound sources that are outside the reach of U.S. federal regulations, such as international shipping. Finally, policy must frame the questions about underwater noise in a way that research can answer. Specifically, how much noise exists, how fast is it growing, where is it, and what consequences does it have for animals and their ecosystems?

9:00

3aNS4. Determination of background sound level for use with Massachusetts community sound level criteria. Gregory Tocci (Cavanaugh Tocci Assoc., Inc., 327F Boston Post Rd., Sudbury, MA 01776)

In the early 1970s, the Commonwealth of Massachusetts began the use of their community sound level criteria. In 1990, these criteria were formalized as Policy 90-001. The policy states that: A source of sound will be considered to be violating the Department's noise regulation (310 CMR 7.10) if the source increases the broadband sound level by more than 10 dB(A) above the ambient. Ambient is defined as the background A-weighted sound level that is exceeded 90% of the time measured during equipment operating hours. The Policy also address tonal sound, and this will be described, but the primary concern of this presentation is the measurement of background sound. A number of background sound measurement strategies will be discussed and a recommendation made on how background should be measured and analyzed that is easily understood by regulators, developers, and the community.

9:20

3aNS5. Noise regulations that serve the community. James D. Barnes (Acentech, Inc., 33 Moulton St., Cambridge, MA 02138)

State and local governments have adopted a variety of regulations over the past 40 years to protect the public against unreasonable noise. These regulations, which may be be qualitative in nature, may prohibit particular activities (e.g., barking dogs or loud radios) or outlaw nuisance noise in general (e.g., sounds disturbing to a resident). In addition, these regulations may be quantitative and define specific sound limits for a source at property line or community locations. The quantitative approach may typically include absolute limits that account for many factors, such as the time of day or week, the type of source and receptor, and the character of the sound; or in the case of Massachusetts, may include relative limits that account for the ambient background sound levels. A significant increase in community sound levels that is caused by a source is a useful indicator of a potential noise problem, but should this factor be considered when setting regulatory limits for individual projects? The presentation shall explore this issue and its effect on the process of siting, permitting, designing, and operating industrial facilities.

9:40

3aNS6. The need to revitalize our national noise policy. William W. Lang (29 Hornbeck Ridge, Poughkeepsie, NY 12603, langww@alum.mit.edu) and Leo L. Beranek (Cambridge, MA 02138)

At the present time, our Nation does not have a national noise policy. The Congress attempted to define one with the passage of the Noise Control Act of 1972 as follows: *The Congress declares that it is the policy of the United States to promote an environment for all Americans free from noise that jeopardizes their health and welfare*. The Act assigned to the Environmental Protection Agency (EPA) the role of leading Federal agency with the task of coordinating the programs of all Federal agencies relating to noise research and noise control. Ten years later (1982), all funds for noise control were withdrawn from the EPA, and today the Nation is without an effective, overall noise policy. Residual responsibility for the control of environmental and occupational noise currently rests with a dozen agencies of the Federal government, as well as State, municipal, and local authorities. But the activities of these organizations are largely uncoordinated, and the enforcement of existing noise control regulations is at best sporadic, and in some cases nonexistent. To assure our quality of life and to minimize the economic impact of potential trade barriers, a study team is recommending that a national noise policy be developed and then implemented in the immediate future.

10:00–10:15 Break

10:15

3aNS7. Revisiting the NIOSH *Criteria for a Recommended Standard: Occupational Noise Exposure*. William J. Murphy and John R. Franks (Natl. Inst. for Occupational Safety and Health, Hearing Loss Prevention Section, 4676 Columbia Pkwy., M.S. C-27, Cincinnati, OH 45226-1998, wjm4@cdc.gov)

In 1998, the National Institute for Occupational Safety and Health (NIOSH) revised the *Criteria for a Recommended Standard: Occupational Noise Exposure* [DHHS (NIOSH) Publication No. 98-126]. NIOSH reevaluated the recommended exposure limit (REL) for occupational noise exposure and reaffirms support for 85-dBA REL. Based upon scientific evidence, NIOSH recommends a 3-dB exchange rate. NIOSH recommends that significant threshold shift be identified as an increase of 15 dB in the hearing threshold level at 500, 1000, 2000, 3000, 4000, or 6000 Hz in either ear, with two consecutive audiometric tests. The new criterion has the advantages of a high identification rate and a low false-positive rate. In contrast with the former 1972 criterion, NIOSH no longer recommends age correction on individual audiograms. NIOSH has revisited its recommendations on the using of single-number laboratory-derived Noise Reduction Rating (NRR) required by the U.S. Environmental Protection Agency for labeling of hearing protectors sold within the United States. In 1972, NIOSH recommended the use of the full NRR value; however, the new criterion recommends derating the NRR by 25%, 50%, and 70% for earmuffs, formable earplugs, and all other earplugs, respectively. This presentation will compare and contrast current regulations against the NIOSH recommendations.

10:35

3aNS8. Benefits and disadvantages of self-regulation of environmental noise from military training. George A. Luz (U.S. Army Ctr. for Health Promotion and Preventive Medicine, 5158 Blackhawk Rd., Aberdeen Proving Ground, MD 21010-5403)

In a 1981 Executive decision, the Administration's Office of Management and Budget (OMB) told the Environmental Protection Agency to end funding of the Office of Noise Abatement and Control (ONAC). This decision, coupled with a specific exemption for military equipment contained in the Noise Control Act of 1972, ensured that the military departments would be self-regulating in regard to noise. This self-regulation for noise stands in contrast to the external regulation of other pollutants, such as air and water emissions. Two possible disadvantages of self-regulation are (1) reduced funding for noise management compared with funding for externally regulated pollutants, and (2) lack of an independent and external set of standards for determining acceptable limits on community noise exposure. Three possible benefits are (1) avoiding the costs of mitigating trivial violations of external standards, (2) maintaining a long-standing policy of preventing noise problems through land use planning, and (3) enabling negotiated solutions between installations and their neighboring communities. The paper ends with an examination of a negotiated solution for a community subjected to noise from the detonation of obsolete ammunition.

10:55

3aNS9. Public policy issues in animal bioacoustics. Ann E. Bowles (Hubbs-Sea World Res. Inst., 2595 Ingraham St., San Diego, CA 92109) and Mardi C. Hastings (Biomed. Eng. Ctr., Ohio State Univ., 1080 Carmack Rd., Columbus, OH 43210)

Control of anthropogenic noise in many terrestrial and underwater environments is crucial for maintaining communication, health, and normal behavior of animals. Noise can be an issue for any species; usually, however, endangered and threatened species and marine mammals are the ones provided legal protection under the Endangered Species Act and/or the Marine Mammal Protection Act. Recent substantiated noise effects in the ocean have received much publicity and generated a more global approach to noise control. However, there are also cases where publicity was not accompanied by scientific data substantiating effects (e.g., an incident involving a $1 million noise barrier installed to protect passerine birds). The public and environmental managers have had difficulty developing adequate guidelines not only because necessary data are often lacking, but also because the manner in which funding is allocated—noise-producing agencies or private organizations are often pressured to fund studies—gives rise to inevitable conflicts of interest (or the perception thereof). Examples of recent noise-related controversies will be presented to examine the role of scientists, engineers, and professional organizations such as ASA in dealing with conflicts of interest and formulating public policy.

11:15

3aNS10. Community-based environmental noise management. Lawrence S. Finegold (Finegold & So, Consultants, 1167 Bournemouth Ct., Centerville, OH 45459, LSFinegold@earthlink.net)

After a series of major successes in implementing national environmental noise policies from the 1960s to the 1980s, the U.S. has not kept up with improvements in environmental noise policies seen in other areas of the world, most notably in Europe during the past decade. The cessation of funding for the EPA Office of Noise Abatement and Control (ONAC) in 1981 was a defining point in the history of U.S. environmental noise policy, leading to a virtual stoppage of the development of new noise regulations. The Noise Control Act (NCA) of 1972, which established EPA ONAC, was never rescinded after ONAC funding was abolished, but it is not being implemented either. Thus national noise policies are generally not being enforced, except for aircraft certification regulations, and help for state and local efforts is inadequate. In this paper we describe an evolving concept, community-based environmental noise management, to provide practical guidance to local communities in developing noise management strategies. This concept includes components such as local noise ordinances, coordination with State and Federal agencies, incentive and cost-sharing programs, land use planning, an updated environmental impact analysis process, and a negotiation and dispute resolution process. Each of these will be briefly discussed.

WEDNESDAY MORNING, 5 JUNE 2002 STERLINGS ROOMS 2 AND 3, 8:00 TO 10:00 A.M.

Session 3aPA

Physical Acoustics: Surface Waves and Ultrasound Spectroscopy

Joseph R. Gladden, Cochair
Physics Department, Pennsylvania State University, 104 Davey Laboratory, University Park, Pennsylvania 16802

Ralph T. Muehleisen, Cochair
Civil, Environmental and Architectural Engineering, University of Colorado, 428 UCB, 441 Ecot, Boulder, Colorado 80309-0428

Contributed Papers

8:00

3aPA1. Acoustically excited surface waves on empty or fluid-filled cylindrical and spherical shells. A. Claude Ahyi, H. Cao, P. K. Raju (Dept. of Mech. Eng., Auburn Univ., Auburn, AL 36849-5341), M. F. Werby (Dept. of Phys., Catholic Univ. of America, Washington, DC 20064-1 and NRL Code 7180, Stennis Space Center, MS 39529), X. L. Bao, and H. Überall (Dept. of Phys., Catholic Univ. of America, Washington, DC 20064-1)

A comparative study is presented of the acoustical excitation of circumferential (surface) waves on fluid-immersed cylindrical or spherical metal shells, which may be either evacuated, or filled with the same or a different fluid. The excited surface waves can manifest themselves by the resonances apparent in the sound scattering amplitude, which they cause upon phase matching following repeated circumnavigations of the target object, or by their re-radiation into the external fluid in the manner of head waves. We plot dispersion curves versus frequency of the surface waves, which for evacuated shells have a generally rising character, while the fluid filling adds an additional set of circumferential waves that descend with frequency. The resonances of these latter waves may also be interpreted as being due to phase matching, but they may alternately be interpreted as constituting the eigenfrequencies of the internal fluid contained in an elastic enclosure.

8:15

3aPA2. Acoustic resonances of fluid-immersed elastic cylinders and spheroids: Theory and experiment. Jan Niemiec (Naval Surface Warfare Ctr., Carderock Div., West Bethesda, MD 20817-5700), Herbert Überall, and X. L. Bao (Catholic Univ. of America, Washington, DC 20064-1)

Frequency resonances in the scattering of acoustic waves from a target object are caused by the phase matching of surface waves repeatedly encircling the object. This is exemplified here by considering elastic finite cylinders and spheroids, and the phase-matching condition provides a means of calculating the complex resonance frequencies of such objects. Tank experiments carried out at Catholic University, or at the University of Le Havre, France by G. Maze and J. Ripoche, have been interpreted

using this approach. The experiments employed sound pulses to measure arrival times, which allowed identification of the surface paths taken by the surface waves, thus giving rise to resonances in the scattering amplitude. A calculation of the resonance frequencies using the T-matrix approach showed satisfactory agreement with the experimental resonance frequencies that were either measured directly (as at Le Havre), or that were obtained by the interpretation of measured arrival times (at Catholic University) using calculated surface wave paths, and the extraction of resonance frequencies therefrom, on the basis of the phase-matching condition. Results for hemispherically endcapped, evacuated steel cylinders obtained in a lake experiment carried out by the NSWC were interpreted in the same fashion.

8:30

3aPA3. Experimental study of the A-wave axial propagation inside a fluid filled cylindrical shell: Application to Gelification process monitoring by surface acoustic wave. Loic Martinez, Brahim Senouci, Stephane Serfati (Laboratoire d'Electronique Applique de Cergy Pontoise, Universite de Cergy Pontoise, 5 mail Gay-Lussac, F 95 031 Neuville sur Oise Cedex, France), Marcel Gindre, Jean-Yves Le Huerou (Laboratoire dImagerie Parametrique, 75270 Paris Cedex 06, France), and Pascal Griesmar (Laboratoire de Chimie des Materiaux Inorganiques, 95 031 Neuville sur Oise Cedex, France)

During gel formation, the material passes from the liquid state to the solidlike state. In order to characterize this liquid-to-solid transition, the low-frequency fluid-born plate mode A-wave is used. The A-wave has interesting properties to be nonattenuated—very dispersive at low-frequency thickness products (fd), and nondispersive for high fd (the Scholte wave is the high fd limit). The axial propagation of the A-wave is studied along a glass cylindrical shell in the mid to high fd range (0.5 MHz mm—1.5 MHz mm). The A-wave is generated and received at one extremity of the tube by a contact broadband transducer excited by a pulse. In the calibration test, the cylindrical shell is filled with water and surrounded by air outside. The A-wave dispersion curve is extracted by short time Fourier transform; comparison with the theoretical dispersion curves is very good. For the gelification monitoring, the water is replaced by the gel in its liquid state. A-wave time signals are recorded during the gelification process. The evolution of the longitudinal speed during the gelification time is deduced from the A-wave dispersion curves.

8:45

3aPA4. Differential surface acoustic wave IDT gyroscope. Jose K.A, Anshu Mehta, and Vijay K. Varadan (212 EES Bldg., Penn State Univ., University Park, PA 16802)

This paper presents the design and experimental evaluation of a gyroscope based on differential design of a surface acoustic wave (SAW) resonator and a sensor on a piezoelectric substrate. The resonator is divided into two halves, one with metallic dot arrays and other without any metallic dots. Standing waves are formed inside both cavities. The particles at the antinodes of the standing wave pattern experience large amplitude of vibration that serves as the reference vibrating motion for this gyroscope. The metallic dots strategically positioned at the antinode locations experience Coriolis force due to rotation and acoustically amplify the magnitude in the orthogonal direction. However, the other half of the resonator cannot generate any secondary SAW. A wideband SAW sensor arranged orthogonal to the SAW resonator picks up these two secondary SAWs and is fed to a differential amplifier. Any drift in the gyroscope signal can be completely eliminated by this novel design. The performance of this 74.2-MHz gyroscope shows very high sensitivity and dynamic range, which is ideal for many commercial applications. Unlike other MEMS gyroscopes, this gyroscope has a planar configuration with no suspended resonating mechanical structures, thereby being inherently robust and shock resistant.

9:00

3aPA5. Diffraction correction for precision SAW velocity measurements. Alberto Ruiz M. and Peter B. Nagy (Dept. of Aerosp. Eng. and Eng. Mech., Univ. of Cincinnati, Cincinnati, OH 45221-0070, ruiza@email.uc.edu)

Surface acoustic wave (SAW) dispersion measurements can be used to nondestructively characterize shot-peened, laser shock-peened, and other surface-treated specimens. In recent years, there have been numerous efforts to separate the contribution of surface roughness from those of near-surface material variations, such as residual stress, texture, and increased dislocation density. This talk addresses the problem that a perceivable dispersive effect, similar to the one found on rough shot-peened specimens, is exhibited by untreated smooth surfaces as well. The dispersion measurements were performed using laser-ultrasonic scanning combined with special digital signal-processing methods. The observed dispersion effect is on the order of 0.1%, which is comparable to the expected velocity change produced by near-surface compressive residual stresses in metals below their yield strength. The cause of this apparent dispersion is the diffraction of the SAW as it travels over the surface of the specimen. It is demonstrated that a diffraction correction may be introduced to increase the accuracy of surface wave dispersion measurements. A simple diffraction correction model was developed for surface waves; this correction was subsequently validated by laser-interferometric surface wave velocity measurements on 2024-T351 aluminum specimens.

9:15

3aPA6. Thin film characterization using resonant ultrasound spectroscopy. J. R. Gladden, Jin H. So, Rajdeep Pradhan, and J. D. Maynard (Dept. of Phys., Penn State Univ., University Park, PA 16803, jgladden@psu.edu)

With mechanical and electrical components approaching nanoscale dimensions, there is great interest in the properties of thin films. Some properties, such as structure and phase transitions, can be probed with measurements of elastic constants. We have been using Resonant Ultrasound Spectroscopy (RUS) to probe such properties for a variety of thin film materials including colossal magnetoresistance (CMR) materials, carbon nanotubes, and a magnetic semiconducting film on GaAs. By accurately characterizing the substrate, the elastic constants and damping of a deposited film can be determined from changes in the frequency and quality factor (Q) of the sample resonances. Issues of precision and accuracy in RUS measurements are important for application to thin films. If one is interested in relative changes in properties (e.g., resulting from temperature variations, phase transitions, etc.), the precision from sharp resonances ($Q > 5{,}000$) is more than adequate. However, since a film may occupy only 1/1000 of the sample, attaining high accuracy is challenging, as uncertainty in sample dimensions and crystallographic orientation can have severe effects on results. Temperature data for the magnetic films mentioned and attenuation of carbon nanotube films will be presented. [Work supported by ONR.]

9:30

3aPA7. Effect of hydrochloric acid on sound absorption and relaxation frequency in magnesium sulfate solutions. F. H. Fisher (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0701)

The epic work of Kurtze and Tamm on sound absorption spectroscopy in divalent sulfate electrolyte solutions (1953) from the low-kHz region up to over 200 MHz revealed astonishing variability at frequencies below 10 MHz and a common relaxation frequency at about 200 MHz. For magnesium sulfate [Epsom salts] solutions, the salt producing 30× the absorption of fresh water below the 100-kHz region in the oceans at low concentrations [$\sim$0.02 moles/liter], Kurtze and Tamm investigated the effects of adding HC1 or H_2SO_4. They found that as formal pH increased, the results were different for these acids in reducing the sound absorption. Fisher (1983) found that if the absorption was plotted against free hydrogen, ion concentration was the same. We used the 100-liter titanium sphere, a spare ballast tank from the WHOI submarine ALVIN. With precise

temperature control, we found an increase in the relaxation frequency as HC1 was added in conjunction with the reduction in sound absorption. The results will be presented and an explanation will be proposed in the context of the Eigen and Tamm multistate dissociation model for $MgSO_4$ (1962) which explains the effects of pressure on both absorption and conductance. [Work supported by ONR.] The author acknowledges C. C. Hsu for his work on this project.

9:45

3aPA8. Using resonance ultrasound spectroscopy to study colossal magnetoresistance in thin films. Rajdeep Pradhan, Jin H. So, J. R. Gladden, and J. D. Maynard (Dept. of Phys., Penn State Univ., University Park, PA 16803, rsp133@psu.edu)

Materials displaying colossal magnetoresistance (CMR) are candidates for important applications such as high density data storage. When the CMR material is in the form of a thin film on a substrate, the strain which results from lattice mismatch with the substrate alters the behavior of the material. As for CMR materials in bulk form, there should be a correlation between the CMR behavior and the structural and elastic properties of the strained thin films. Measurements of the elastic properties of thin films are obtained using a novel small sample resonant ultrasound spectroscopy (RUS) technique. Thin films of the magnetic material are deposited on a substrate and the difference in natural frequencies are then used to obtain the elastic constants of the film. The dominant CMR materials are perovskite manganites, a transition metal oxide. They exhibit a variety of exotic structural, magnetic and electrical behaviors, which is yet to be understood. In this paper measurements of the electrical resistance and elastic constants of 200 and 400 nm thin films, as well as unstrained samples, of the CMR material $La(0.7)Ca(0.3)MnO_3$ are presented. [Work supported by ONR.]

WEDNESDAY MORNING, 5 JUNE 2002 GRAND BALLROOM 4, 8:15 TO 11:50 A.M.

Session 3aPP

Psychological and Physiological Acoustics and Speech Communication: Honoring the Contributions of Ira J. Hirsh

Janet M. Weisenberger, Cochair
College of Social and Behavioral Science, The Ohio State University, 1010 Derby Hall, 154 North Oval Mall, Columbus, Ohio 43210

Judith L. Lauter, Cochair
Department of Human Services, Stephen F. Austin University, P.O. Box 13019, SFA Station, Nacogdoches, Texas 75962

Invited Papers

8:15

3aPP1. Ira at 80: The acronyms of a career in acoustics. Janet M. Weisenberger (Speech and Hearing Sci., Ohio State Univ., Columbus, OH 43210, jan+@osu.edu)

In a career that spans some 54 years to date, the name of Ira J. Hirsh has been associated with significant scientific contributions to psychoacoustics, outstanding mentoring of research scientists, and dedicated service to the fields of acoustics, audiology, and psychology. It is a career that can be traced by acronyms that are part of the daily vocabulary of hearing scientists. These include acronyms of location: Early work at the Psychoacoustics Laboratory at Harvard (PAL), a long tenure in research at the Central Institute for the Deaf (CID), service as faculty member, chair, and dean at Washington University (WashU); acronyms of professional societies that have honored him: Acoustical Society of America (ASA), International Commission of Acoustics (ICA), American Psychological Association (APA), American Psychological Society (APS), American Speech-Language-Hearing Association (ASHA), American Association for the Advancement of Science (AAAS); acronyms of his service to the National Academy of Science: National Research Council (NRC), Commission on Behavioral and Social Science and Education (CBASSE); and acronyms of his contributions to psychoacoustics: Masking Level Difference (MLD), Temporal Order Judgments (TOJ). In large part, these acronyms are part of our vocabulary because of Ira's contributions, and tracing them over the past half-century yields a substantive look at the development of the field of hearing science.

8:30

3aPP2. Events, sequences, and patterns: Hirsh's prescient proposals. Charles S. Watson (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405, watson@indiana.edu)

Over four decades ago, Ira Hirsh was one of the first to recognize the need for auditory research to expand beyond the study of single tones, noise burst, and clicks. He wrote, ''We propose to examine auditory perception at a more complex level. The discrimination among and identification of single sounds, which we shall refer to as acoustic *events*, is undoubtedly an important part of the perceptual process, but perhaps more important are the rules by which we distinguish and identify sequences of acoustic events. Auditory psychophysics has been concerned with the acoustic characteristics of the sequence parts, the events; but we need to know more about the ways in which the parts combine to form patterns which, since they are generated in time, we may call *sequences*'' [J.

Acoust. Soc. Am. **31** (1959)]. This must be one of the more prescient proposals of that era, followed as it was not only by Hirsh's series of studies of temporal processing, but by a raft of other adventures into the world he forecast, including research on a great variety of auditory patterns and sequences. Theoretical interpretations of that work and of the general problem areas that Hirsh pioneered will be discussed.

8:50

3aPP3. It's about time: Presentation in honor of Ira Hirsh. Ken Grant (Walter Reed Army Medical Ctr., Army Audiol. and Speech Ctr., Washington, DC 20307-5001)

Over his long and illustrious career, Ira Hirsh has returned time and time again to his interest in the temporal aspects of pattern perception. Although Hirsh has studied and published articles and books pertaining to many aspects of the auditory system, such as sound conduction in the ear, cochlear mechanics, masking, auditory localization, psychoacoustic behavior in animals, speech perception, medical and audiological applications, coupling between psychophysics and physiology, and ecological acoustics, it is his work on auditory timing of simple and complex rhythmic patterns, the backbone of speech and music, that are at the heart of his more recent work. Here, we will focus on several aspects of temporal processing of simple and complex signals, both within and across sensory systems. Data will be reviewed on temporal order judgments of simple tones, and simultaneity judgments and intelligibility of unimodal and bimodal complex stimuli where stimulus components are presented either synchronously or asynchronously. Differences in the symmetry and shape of ''temporal windows'' derived from these data sets will be highlighted.

9:10

3aPP4. The times of Ira Hirsh. Pierre L. Divenyi (Speech and Hearing Res., VA Medical Ctr., Martinez, CA 94553)

Ira Hirsh was among the first to recognize that the auditory system does not deal with temporal information in a unitary way across the continuum of time intervals involved in speech processing. He identified the ''short'' range (extending from 1 to 20 ms) as that of phase perception, the ''medium'' range between 20 and 100 ms as that in which auditory patterns emerge, and the ''long'' range from 100 ms on as that of separate auditory events. Further, he was also among the first to recognize that auditory time perception heavily depended on the spectral context. A study of the perception of sequences representing different temporal orders of three tones, by Ira and the author [e.g., Divenyi and Hirsh, J. Acoust. Soc. Am. **64**, 1369–1385 (1978)] demonstrated the dependence of auditory sequence perception on both time range and spectral context, and provided a bridge between Hirsh's view of auditory time and Bregman's view of stream segregation [*Auditory Scene Analysis* (MIT Press, Cambridge, MA, 1991)]. A subsequent search by the author for psychophysical underpinnings of the cocktail-party phenomenon [e.g., Divenyi *et al., Moh. Mtn. Workshop Appl. SPAA* (IEEE, Bellingham, WA, 1997)] suggest that the segregation of simultaneous streams of speech may rely on the ability to follow spectral changes in the demisyllabic-to-syllabic (100- to 200-ms) range, i.e., Ira's ''long'' range.

9:30

3aPP5. Ira as a pioneer in audiology: His contributions to the clinical measurement of hearing and hearing impairment. C. Formby (Univ. of Maryland, Baltimore, MD 21201) and J. P. Gagne (Univ. de Montreal, Montreal, QC H3C 3J7, Canada)

Ira Hirsh's contributions to clinical science and research are diverse and significant. In fact, approximately one-third of the 100+ publications that Ira lists in his *curriculum vitae* (CV) are clinical in nature, dealing with various aspects of audiology, deafness, hearing aids, aural rehabilitation, and speech and language pathology. The majority of these citations, fully one-quarter of his publication list, addresses problems specific to the clinical measurement of hearing and hearing impairment. Undoubtedly, the most influential of these published works appears in his CV under the citation ''The Measurement of Hearing.'' The forward for this publication, his only textbook, was penned in June, 1952 (now precisely half a century past at the time of this session). The aims of this presentation are to (1) provide perspective on the fundamental importance of his virtually timeless text in shaping the fledgling discipline of audiology, and (2) celebrate Ira's many contributions to the profession and practice of audiology. [Preparation for this presentation was supported, in part, by a K24 career development award from NIDCD.]

9:50–10:10 Break

10:10

3aPP6. Ira Hirsh and oral deaf education: The role of audition in language development. Ann Geers (Central Inst. for the Deaf, 4560 Clayton Ave., St. Louis, MO 63110)

Prior to the 1960s, the teaching of speech to deaf children consisted primarily of instruction in lip reading and tactile perception accompanied by imitative exercises in speech sound production. Hirsh came to Central Institute for the Deaf with an interest in discovering the auditory capabilities of normal-hearing listeners. This interest led him to speculate that more normal speech development could be encouraged in deaf children by maximizing use of their limited residual hearing. Following the tradition of Max Goldstein, Edith Whetnall, and Dennis Fry, Hirsh gave scientific validity to the use of amplified speech as the primary avenue to oral language development in prelingually deaf children. This ''auditory approach,'' combined with an emphasis on early intervention, formed the basis for auditory-oral education as we know it today. This presentation will examine how the speech perception, language, and reading skills of prelingually deaf children have changed as a result of improvements in auditory technology that have occurred over the past 30 years. Current data from children using cochlear implants will be compared with data collected earlier from children with profound hearing loss who used hearing aids. [Work supported by NIH.]

10:30

3aPP7. Personal glimpses of Ira Hirsh: Covariance of perception and reality. William Clark (Central Inst. for the Deaf, 4560 Clayton Ave., St. Louis, MO 63110, wclark@cid.wustl.edu)

No session honoring Ira Hirsh would be complete without a personal reflection of his enormous influence on his colleagues, students, and friends. During his career in St. Louis that covered the entire second half of the 20th Century (and still continues), Ira interacted with numerous colleagues at Central Institute for the Deaf, where he served as a senior scientist, Director of Research, and Institute Director. His second parallel career was at Washington University, where he was Professor of Psychology, and included a term as Dean of the Faculty of Arts and Sciences. This tribute will provide reflections and reminiscences of what it was like to work with, and learn from, Ira Hirsh.

10:50

3aPP8. Parameter space of the dichotic speech experiment. Robert C. Bilger (Dept. of Speech & Hearing Sci., Univ. of Illinois at Urbana-Champaign, 901 S. 6th St., Champaign, IL 61820, r-bilger@uiuc.edu) and Charles E. Speaks (Univ. of Minnesota, Minneapolis, MN 55455)

The experiment in which different CV nonsense syllables are presented simultaneously to each ear and the listener is asked to identify both consonants often results in more correct responses to syllables presented to the right than to the left ear, the so-called right-ear advantage (REA). The REA is generally interpreted as an indication that the left hemisphere processes speech more efficiently than does the right hemisphere. As the experiment is usually conducted, however, it has the parameter space of 3 (i.e., requires three numbers to completely specify the results). Using data obtained independently in each of our labs, it will be shown that the three parameters necessary to uniquely specify the results of the experiment are the probability that the (1) first response is correct; (2) first response is to the syllable presented to the right ear; and (3) second response will be incorrect given that the first response was correct. The ear advantage (R, 0, or L) is predicted to the nearest percent using the first two parameters; but, as of this moment, the exact fourfold matrix for a listener cannot be predicted.

11:10

3aPP9. Transcending boundaries with Ira Hirsh. Punita G. Singh (Sound Sense, C6-6046 Vasant Kunj, New Delhi 110070, India, soundsense@vsnl.com)

Ira Hirsh has made many contributions to various fields of acoustics from speech, hearing, psychological and physiological acoustics, to musical and architectural acoustics. It was a privilege for me to have been his student in all these areas, and to have had him as a guide through masters and doctoral degree programs that focused on topics that lie at the boundaries connecting these disciplines. Ira was not a prescriptive advisor, imposing particular research topics or procedures on his graduate students. Rather, he encouraged originality, innovation, and personal goal setting. He would subtly suggest starting points and provide landmarks as references, rather than explicit directions leading to them. One had to navigate the path by ones own wits. This approach encouraged lateral, out-of-the box thinking, while also leading to respectful appreciation of historic trajectories in scientific research. During our time together, we worked on several aspects of music, including, rhythm, melody, pitch, and timber perception. Some of this work will be recapitulated, highlighting Ira's role in its exposition and development. His multidimensional personality, astute insights, colorful remarks, wry humor, care, and concern are qualities to be cherished—beyond the boundaries of campus, city, country, and contemporaneity.

11:30

3aPP10. Ad cerebrum per scientia: Ira Hirsh, psychoacoustics, and new approaches to understanding the human brain. Judith Lauter (Human Neurosci. Lab., Dept. of Human Services, Stephen F. Austin State Univ., Box 13019, SFA Station, Nacogdoches, TX 75962)

As Research Director of CID, Ira emphasized the importance of combining information from biology with rigorous studies of behavior, such as psychophysics, to better understand how the brain and body accomplish the goals of everyday life. In line with this philosophy, my doctoral dissertation sought to explain brain functional asymmetries (studied with dichotic listening) in terms of the physical dimensions of a library of test sounds designed to represent a speech–music continuum. Results highlighted individual differences plus similarities in terms of patterns of relative ear advantages, suggesting an organizational basis for brain asymmetries depending on physical dimensions of stimulus and gesture with analogs in auditory, visual, somatosensory, and motor systems. My subsequent work has employed a number of noninvasive methods (OAEs, EPs, qEEG, PET, MRI) to explore the neurobiological bases of individual differences in general and functional asymmetries in particular. This research has led to (1) the AXS test battery for assessing the neurobiology of human sensory-motor function; (2) the handshaking model of brain function, describing dynamic relations along all three body/brain axes; (3) the four-domain EPIC model of functional asymmetries; and (4) the trimodal brain, a new model of individual differences based on psychoimmunoneuroendocrinology.

Session 3aSAa

Structural Acoustics and Vibration: Scattering; Composite Materials

Joseph W. Dickey, Chair
Center for Nondestructive Evaluation, The Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218

Contributed Papers

8:00

3aSAa1. Chaos, catastrophe, and critical frequency points on elastic shells. Michael Werby (NRL, Code 7181, Stennis Space Center, MS 39529) and H. Uberall (Dept. of Phys., Catholic Univ. of America, Washington, DC)

Exact equations exist for the determination of the backscattered signals excited by acoustic signals on submerged elastic shells. In the absence of resonances these signals may be cast in terms of conservation principles of inertial components of the interaction of the signal with the object. This indeed leads to an adequate description of the acoustic background of such targets. What is determined to be critical is the inclusion of an entrained mass related to the displacement of the mass equivalent of the fluid due to the movement of the object. But what happens to such inertial effects in the presence of elastic resonances? We explore this issue and demonstrate that the effective or inertial component ranges from zero to infinity over a small frequency range that characterizes the resonance width. Over this range there is also a phase shift of 1800. We may as well examine the effect of radiation loading in this range. It is also possible to investigate a phase plot of the real and imaginary components of the effective mass factor which leads to closed trajectories for certain classes of resonances and to hyperbolic trajectories for other types. The meaning of these observations is discussed. [Work supported by NRL and ONR.]

8:15

3aSAa2. Backscattered acoustic signal from submerged water filled elastic shells can be decomposed into body resonances and eigenmodes of the contained water. Michael Werby (NRL Code 7180, Stennis Space Center, MS 39529) and H. Uberall (Dept. of Phys., The Catholic Univ. of America, Washington, DC 20064)

Acoustic signals scatter from an elastic shell and excite elastic shell resonances. For evacuated shells resonances due to proper Lamb waves, A0 and S0 waves and the pseudo-Stoneley resonances, are the only ones allowed. When the shell is fluid filled then the eigenfrequencies of the included fluid may be excited. Further, the presence of the fluid may alter the elastic body resonances. When the impedance of the entrained fluid is not much smaller than the mechanical impedances of the elastic material then the scattered signal is greatly influenced by the entrained fluid, but only at their allowed eigenfrequencies. Since the fluid is entrained this leads to a discrete spectrum and an eigenvalue problem which is quite manageable. In this work we outline a method for determining the eigenfrequencies as well as their nature and use their values to isolate the actual body resonances due to the elastic material. We also illustrate the influence of the presence of the entrained fluid on the elastic resonance frequencies. [Work supported by NRL and ONR.]

8:30

3aSAa3. Visualization of the energy flow for a guided forward wave in and around a fluid loaded elastic cylindrical shell: Color coding of the Poynting vector field. Cleon E. Dean (Phys. Dept., P.O. Box 8031, Georgia Southern Univ., Statesboro, GA 30460-8031) and James P. Braselton (Dept. of Mathematics and Computer Sci., Georgia Southern Univ., P.O. Box 8093, Statesboro, GA 30460-8093)

Previous work [Cleon E. Dean and James P. Braselton, ''Visualization of the energy flow for a guided forward wave in and around a fluid loaded elastic cylindrical shell,'' J. Acoust. Soc. Am. **109**, 2379 (2001)] showed the energy flow for a forward propagating fluid loaded elastic cylindrical shell at the resonance frequency. The results were difficult to interpret since although the two counterpropagating guided waves were separated and displayed independently, the conventional grid of vector arrows was hard to interpret in part because they were hard to see individually. The current work uses color encoding to show the Poynting vector field. Hue is used to indicate the direction of the energy flow while intensity of the color is used to denote the magnitude of the Poynting vector at that point.

8:45

3aSAa4. Investigation of the sound transmission behavior of a chamber core cylinder. Deyu Li and Jeffrey S. Vipperman (Dept. of Mech. Eng., Univ. of Pittsburgh, 531 Benedum Hall, Pittsburgh, PA 15228)

Several kinds of novel composite structures, such as advanced grid stiffened (AGS) and chamber core (CC) structures have been designed, fabricated, and investigated for both civil and military applications. The chamber core composite is a novel advanced sandwich-type structure that is created by filament winding an inner shell onto a cylindrical mandrel, arranging previously fabricated U-shaped channels around the perimeter of this shell to form the inner chamber walls, and filament winding an outer shell followed by a co-cure process. In this study, the structural/acoustic behavior of a normal composite chamber core cylinder is investigated both theoretically and experimentally. Lightly coupled structural and acoustic modal parameters are identified using experimental modal analysis techniques. The properties of sound transmission loss (TL) of the cylinder are also investigated experimentally. The effect of the structural/acoustic natural frequencies and the damping on the sound transmission loss is analyzed. Finally, passive control strategies are discussed, and several passive control materials for improving the sound transmission loss (0–500 Hz) of the cylinder are experimentally evaluated. [Work sponsored by the Air Force Research Laboratory Space Vehicles Directorate (AFRL/VS), POC Dr. Steven Lane, (505) 846-9944.]

9:00

3aSAa5. Sound absorption characteristics of cenosphere enriched cement and asphalt concrete. Vikrant Tiwari, Arun Shukla (Dynamic Photomechanics Lab., Dept. of Mech. Eng. and Appl. Mech., Univ. of Rhode Island, Kingston, RI 02881), and Arjit Bose (Univ. of Rhode Island, Kingston, RI 02881)

A detailed experimental study has been conducted to determine the effect of addition of cenospheres on the acoustic properties of cement matrix and asphalt concrete. Cenospheres are hollow ceramic microballoons, being a waste product of thermal power plants, they are relatively

inexpensive and their use has the added benefit of decreasing the strain on the environment. Specimens were made with a different volume fraction of cenospheres (0% to 70% in increments of 10) to observe the change in acoustic properties with an increase in the cenosphere content. Specimens of different sizes were tested to study the acoustic characteristics of the materials for a wide range (125 to 4000 Hz) of frequencies. The effect of an increase in the cenosphere content on the dilatational wave velocity in the specimens was also studied with the help of acoustic transducers. It was noted that, although the wave velocity remains relatively unaffected with a change in the cenosphere content, the addition of cenospheres improves the sound absorption characteristics of the cement matrix and asphalt concrete with a considerable decrease (up to 41%) in the density of the material. It has also been noted that the sound absorption of cement increases with the addition of cenosphere up to 40% volume fraction and starts decreasing with the further addition of cenospheres.

9:15–9:30 Break

9:30

3aSAa6. An equivalent solid (u) formulation for poroelastic materials. Dominic Pilon and Raymond Panneton (GAUS, Dept. of Mech. Eng., Univ. de Sherbrooke, 2500 boul. de l'Universite, Sherbrooke QC J1K 2R1, Canada)

Finite element formulations based on Biot's poroelasticity equations have been used extensively throughout recent years. The most common are Biot's displacement ($\underline{u}$,$\underline{U}$) and mixed displacement-pressure ($\underline{u}$,p) formulations. They are used to predict the structural and acoustical behavior of poroelastic mediums within multilayered structures. These models, while accurate, lead to the resolution of large linear systems: they need, respectively, six and four degrees of freedom per node in order to efficiently describe the poroelastic medium's vibroacoustic behavior in 3-D problems. In this paper, a simplified displacement ($\underline{u}$) formulation is presented. It is also based on Biot's equations, but requires only three degrees of freedom per node, related to the solid phase displacement field, to describe the behavior of poroelastic mediums in 3-D problems. The development of the governing ($\underline{u}$) equations for the equivalent solid poroelastic formulation is detailed. The limitations of this approach are also discussed. Numerical and experimental validations are presented in order to show the accuracy and effectiveness of the formulation within its prescribed field of application.

10:00

3aSAa7. Genetic encoding for sound package configuration optimization. Heng-Yi Lai (Dept. of System and Network, United Technologies Res. Ctr., 411 Silver Ln., MS129-17, East Hartford, CT 06108, laih@utrc.utc.com)

The acoustic properties of poroelastic materials can be accurately predicted by the Biot–Allard model if the required acoustic parameters have been successfully characterized. In conjunction with the transfer matrix approach, one can use the model to analytically design an acoustic liner comprising different materials that can provide better performance or cost reduction advantages over their homogeneous counterparts. However, optimizing the layer configuration remains to a challenging task as it is a combinatorial problem and the search for the optimal configuration can be difficult or expensive if simply based on the build-and-test approach. The Generic Algorithm (GA) was identified as a better optimization method over the gradient-based methods for this task due to its nature in finding the combinatorial solutions. This approach is, to the best of our knowledge, novel in applying the GA to this design problem. The performance of the GA method is compared with the exhaustive search in maximizing the sound absorption coefficient of an acoustic liner with constraints set for its cost, thickness, and total weight. The features of the Generic Algorithm and the preliminary proof-of-concept results will be discussed and presented.

10:15

3aSAa8. On the dynamics of fuzzy structures. Jean-Mathieu Mencik and Alain Berry (GAUS, Mech. Eng. Dept., Univ. of Sherbrooke, Sherbrooke, QC J1K 2R1, Canada, Jean-Mathieu.Mencik@gaus.gme.usherb.ca)

In this presentation, one proposes a prediction at low- and mid-frequencies of the dynamics of fuzzy structures. As introduced by Soize, the term ''fuzzy structure'' designates a master structure, whose geometrical, material characteristics, boundary conditions, and excitations are known, coupled with complex systems, called the structural fuzzy or fuzzy, whose characteristics are imprecisely known. Previous works done on this subject analyze the concept of a master structure coupled with a locally homogeneous fuzzy, composed of a large number of linear oscillators excited by their supports. In the present work, the concept of a homogeneous fuzzy is extended to an elastic continuum medium. One theoretically formulates the action of the elastic fuzzy on the master structure: it is shown that the proposed formulation is different from the solution proposed by Soize, derived from the model of a linear oscillator excited by its support. The proposed theory is successfully applied to the case of a homogeneous structural fuzzy composed of a large number of elastic bars whose lengths and cross sections are randomly chosen.

10:30

3aSAa9. Dissipation induced by substructures and distribution of vibratory energy in a complex system. Sunghoon Choi and Yang-Hann Kim (Dept. of Mech. Eng., KAIST, Science Town, Taejon 305-701, Republic of Korea)

This work is concerned with the problem of energy transfer that takes place between a master structure and the substructures attached to it. The response of the system is characterized by the impedance of the substructures and it determines whether the induced damping is real or apparent. When the dissipation is real the master structure has a larger loss factor than that of the substructures and there will be continuous transfer of vibratory energy from the master structure to substructures. However, in the case of apparent damping, one can observe that the vibration energy is transferred back and forth between the master and substructures. Some combinations of the master and substructures have been considered to examine this phenomenon and to determine the criteria for damping. It has found that a modal overlapping condition, which corresponds to bandwidths that exceed the spacing of those natural frequencies, is crucial in determining the characteristics of the system damping. The result of this paper is consistent with that found with the fuzzy structure and SEA framework. [Work sponsored by Ministry of Education, Korean Government under the BK21 program and Ministry of Science and Tech., Korean Government under National Research Lab. program.]

Session 3aSAb

Structural Acoustics and Vibration and Archives and History: History of Structural Acoustics and Vibration

Joseph W. Dickey, Chair
Center for Nondestructive Evaluation, The Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218

Invited Paper

11:00

3aSAb1. History of structural acoustics and vibrations in the Acoustical Society of America. David Feit, Murray Strasberg (Carderock Div., Naval Surface Warfare Ctr., 9500 MacArthur Blvd., West Bethesda, MD 20817), and Eric E. Ungar (Acentech, Inc., Cambridge, MA 02138-1118)

Structural acoustics refers to the interaction of sound and structures—the response of structures to sound, the radiation of sound from vibrating structures, and the effect of the acoustic medium on the structural vibrations. Interest in these subjects increased greatly during the 1930s and 40s because of practical applications in the design of microphones and loud speakers used in telephones, radios, and electronic phonographs. The combination of electrical and mechanical systems lead to the use of electrical engineering concepts such as impedance, circuits, and electrical analogies, in the analysis of mechanical systems. In later years, much of the work dealt with various aspects of underwater structures, prompted by U.S. Navy interests. The field, which began with classical analytical mechanics applications, has progressed to new approaches, including statistical energy analysis, near-field acoustical holography, fuzzy structures, active control of vibrations, and smart materials. In recognition of these new developments, the name of the technical committee was changed in 1987 from ''Shock and Vibration'' to ''Structural Acoustics and Vibration.''

Session 3aUWa

Underwater Acoustics: Array Performance

Peter Gerstoft, Chair
Marine Physics Laboratory, Scripps Institution of Oceanography, University of California, San Diego, 8820 Shellback Way, La Jolla, California 92093-0238

Chair's Introduction—7:55

Contributed Papers

8:00

3aUWa1. Acoustic tracking of towed-array hydrophone positions during tow-vessel maneuvers. Alec J. Duncan (Ctr. for Marine Sci. and Technol., Curtin Univ., GPO Box U1987, Perth, WA 6845, Australia, a.duncan@cmst.curtin.edu.au), Darryl McMahon (Defence Sci. and Technol. Organisation, HMAS Stirling, Garden Island, Western Australia), and Alessandro Ghiotto (Nautronix Ltd., 108 Marine Terrace, Fremantle, WA 6160, Australia)

This paper describes an experiment in which a vessel towing an array of hydrophones executed a series of maneuvers designed to bring the array into orientations favorable for imaging noise sources on the tow-vessel. Signals from tracking beacons on the tow-vessel and transient signals from implosive sources deployed from a second vessel were received on the array hydrophones and were used to obtain independent estimates of hydrophone position. Array shape estimation and tracking algorithms are described and the results of applying these algorithms to the experimental data are presented.

8:15

3aUWa2. Adaptive beamforming of a towed array during a turn. Peter Gerstoft, William S. Hodgkiss, W. A. Kuperman, Heechun Song (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238), Martin Siderius, and Peter L. Nielsen (SACLANT Undersea Res. Ctr., 19138 La Spezia, Italy)

During maneuvering, towed-array beamforming will degrade if a straight array is assumed. This is especially true for high-resolution adaptive beamforming. This problem can be reduced if a proper curved array is assumed. The tow vessel's Global Positioning System (GPS) can be used to estimate this curvature, and this reduces the need for instrumentation in the array. Based on estimated array shape from GPS, both the conventional beamformer and the white-noise-constrained (WNC) adaptive beamformer are shown to track the source well. For calculating the steering vector in the WNC approach, a matrix inversion of the cross-spectral density matrix is involved. This matrix inversion can be stabilized by averaging the cross-spectral density matrix over neighboring frequencies.

8:30

3aUWa3. Subarray partitions of large aperture planar towed arrays. Jennifer A. Watson (Lincoln Lab., MIT, 244 Wood St., Lexington, MA 02420, jwatson@LL.mit.edu), Arthur B. Baggeroer (MIT, Cambridge, MA 02139), Lisa M. Zurk, and Brian H. Tracey (Lincoln Lab., MIT, Lexington, MA 02420)

The current focus of passive detection and localization is in littoral regions where acoustic propagation becomes complicated by severe bottom interaction. The resultant high-transmission loss motivates the need for high-array gain for effective performance over long ranges. Large planar seismic arrays, with aperture dimensions upwards of 3 km$\times$0.5 km, have potential to achieve high gain and good resolution when using matched field processing. In realistic environments, however, large arrays are suceptible to signal gain degradation mechanisms, particularly due to spatial decorrelation of the signal and non-stationary environments. One approach to overcoming this is partitioning the array. Subarray processing reduces stationarity requirements and extracts optimum coherent gain, thus achieving higher gain than that of smaller arrays. This work examines criteria for partitioning planar arrays to perform localization using MFP. Trade-offs between spatial resolution, array gain, and resilience to motion will be quantified and discussed. Performance of different subarray geometries will be presented using adaptive and conventional MFP. [Work sponsored by DARPA under Air Force Contract No. F1962800-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the Department of Defense. AB's work was supported by ONR through the SECNAV CNO Chair.]

8:45

3aUWa4. Passive localization algorithms. Edmund J. Sullivan 46 Lawton Brook Ln., Portsmouth, RI 02871)

An acoustic source can be localized by exploiting the fact that, if the range is not too great, the wavefront of the incoming acoustic energy can be assumed to be circular. This is sometimes called ''Wavefront Curvature Ranging.'' Here, the estimation of range reduces to the estimation of the radius of a circle. In fact, if sufficient horizontal and vertical aperture is available, range, bearing and elevation can be estimated, given a minimum of four receivers. The technique is based on the measurement of the relative phases between pairs of receivers. In this paper, two methods of improving the performance of such localizers are investigated. In the case where the receivers are mounted on a moving platform, it is shown that exploitation of this motion can improve performance by utilizing information that is usually ignored in phase measurements. Also, since the motion introduces a time dependence in the true values of the time delays, a limit on the accuracy of their measurement is incurred. A recursive estimation process is proposed as a remedy for this problem. These techniques will be demonstrated by using simulated data.

9:00

3aUWa5. Beam intensity striations and applications. T. C. Yang (Naval Res. Lab., Washington, DC 20375)

Modal interferences, or striations, are analyzed and modeled for the beam outputs of a horizontal line array obtained by conventional beamforming. The enhancement of signal levels by the array gain (over that of a single element) makes the measurement of the invariant parameter more reliable even in the weak signal cases. It is shown that the signal beams exhibit the same striation pattern as that predicted (theoretically) for a single element. Specifically, for a broadside signal, the beam striation is identical to that of the single element plus a constant signal gain. For a non-broadside target, the signal beam intensity will be modified by a frequency-bearing dependent signal gain due to the signal spread over multiple beams (the multipath effect). Nonetheless the signal beam intensity still retains the same striation pattern (slope) as for a single element. At non-signal look directions, the sidelobe beams (those outside the canonical cones containing the signal arrivals) exhibit an entirely different striation pattern. The difference of signal striations between a fast range-rate, close range signal and distant, slow range-rate interferences can be used to suppress the interferences in array processing. [Work supported by ONR.]

9:15

3aUWa6. Preliminary results of horizontal array coherence from the 2001 ASIAEX South China Sea experiment. Theodore H. Schroeder, James F. Lynch, and Arthur Newhall (Woods Hole Oceanogr. Inst., MS#11, Woods Hole, MA 02543, tschroeder@whoi.edu)

The 2001 ASIAEX (Asian Seas International Acoustics Experiment) South China Sea experiment was performed in April–May, 2001 off the southern coast of the People's Republic of China. This experiment was a combined acoustics/physical oceanography/marine geology effort that endeavored to measure basic acoustic propagation and scattering processes in a well-supported manner. As part of the experiment, a combination vertical/horizontal autonomous receiving array was deployed in 124 m of water, where it listened to a combination of moored and towed acoustic sources spanning a bandwidth of 50–600 Hz. These acoustic measurements were supported by geological surveys and extensive oceanographic measurements from fixed and towed instruments, providing a 3-D picture of the overall oceanography in the experimental area. The initial interest in this data centers about horizontal array coherence issues, as few experiments have had a horizontal array coupled with extensive environmental support, allowing for a direct correlation of the array coherence with the environment. The frequency dependence, time variability, and along-shelf vs. across-shelf variation of coherence are of particular interest. The latest results in these topics will be shown. [Work supported by ONR.]

9:30

3aUWa7. Effect of a ship's body on the beam pattern of its hydroacoustic array. Tran H. Dat and V. T. Grinchenko (Inst. of Hydromech., Natl. Acad. of Sci. of Ukraine, Zhelyabov Str. 8/4, Kiev 03035, Ukraine)

Session 3aUWb

Underwater Acoustics: Shallow Water Propagation and Signal Processing

David M. Deveau, Chair
PSC 1012, Box 701, FPO, AA 34058-9998

Chair's Introduction—10:10

Contributed Papers

10:15

3aUWb1. Shallow water impact acoustics. David M. Deveau (Naval Undersea Warfare Ctr. Det. AUTEC, 801 Clematis St., West Palm Beach, FL 33401)

During a series of impact detection and tracking accuracy tests near Wallops Island, Maryland, an analysis was conducted on the acoustic impacts themselves. Since the impact area was very shallow (60 ft), the acoustic signature traveled unscathed between the 1000 yard baseline sensors. The initial goal was to implement deep water impact tracking algorithms in the shallow environment. Initially the in-water tracking system computed positions that were close to those which were laser sighted, but not close enough. Since the transient acoustic signature was very evident, hence a stable timing mark, an investigation into the acoustics surrounding the impact was undertaken. A review of the data showed small transients which occurred distinctly before the large transient detection that had routinely been considered as the impact itself. Using a threshold detection scheme, a relative set of timing marks were generated based on these early transients and were found to produce a more accurate placement.

10:30

3aUWb2. Channel characterization for communications in very shallow water. Robert J. McDonald, Kerry W. Commander, John S. Stroud, Jo Ellen Wilbur (Coastal Systems Station, 6703 W. Hwy. 98, Panama City, FL 32407-7001, commanderkw@ncsc.navy.mil), and Grant B. Deane (Univ. of California, San Diego, La Jolla, CA 92093-0238)

Broadband acoustic transmissions (7 kHz to 17 kHz) taken from July 4, 2000 to July 8, 2000, in the shallow water near Scripps Pier at La Jolla, California, are used to extract the time-varying channel parameters of coherence time and multipath time delay spread, as functions of frequency and the environmental variables of wave height and tidal fluctuations. Tidal fluctuations, which have a significant effect on water depth at the receiver and transmitter, are shown to strongly correlate to variations in the multipath delay spread. Variations in the coherence time for the channel, as measured by the drop off in correlation between initial and successive impulse responses, are shown to be inversely related to variations in the measured wave height. The coherence time of the channel was found to decrease with increasing center frequency. A scatter plot of the receiver signal-to-noise ratio as a function of wave height and water depth indicates when the shallow water environment allows a viable communications channel to exist.

10:45

3aUWb3. Estimating the velocity of a moving object submerged in an ocean waveguide with active sonar. Yi-san Lai and Nicholas Makris (MIT, 77 Masschusetts Ave., Cambridge, MA 02139)

Standard active sonar and radar systems are often used to estimate the velocity of a moving target in free space by resolving the Doppler shift of the scattered waveform. In an ocean waveguide, multimodal propagation and dispersion make the Doppler effects far more complicated than in free space. In a waveguide, multiple frequency components are typically present in the field scattered from a moving object even if the active source of radiation is harmonic. Applying a free space Doppler correction to the field scattered from a moving object in a waveguide is insufficient to account for the Doppler distortion of the signal in typical ocean waveguides and active sonar scenarios. Spectral and modal formulations for the Doppler-shifted field scattered from a horizontally moving target in a horizontally stratified medium have been developed by Lai and Makris [J. Acoust. Soc. Am. **110**, 2724 (2001)]. These formulations are used with the maximum likelihood method to estimate the velocity of the moving target by modeling the multiply Doppler distorted scattered signal measured with noise in a shallow water waveguide. The performance of the maximum likelihood estimate is evaluated by asymptotic analysis of its bias and mean-square error as a function of signal-to-noise ratio.

11:00

3aUWb4. Time-frequency analysis and conditional moments of shallow-water sound propagation. Patrick J. Loughlin (Dept. of Elec. Eng., Univ. of Pittsburgh, Pittsburgh, PA 15261)

Shallow-water sound propagation is inherently nonstationary owing to such effects as geometric dispersion. The spectrogram has been a principal means to study the nonstationarities and dispersion characteristics of shallow-water sound propagation. In this talk, we give the low-order conditional spectral and wave number moments of sound propagation in dispersive environments, and show how they characterize the nonstationarities induced by channel dispersion on the propagating wave. [Work supported by ONR (N00014-02-1-0084).]

11:15

3aUWb5. Multipath in synthetic aperture sonar. John E. Piper (Coastal Systems Station, 6703 W. Hwy. 98, Panama City, FL 32407)

Synthetic aperture sonar is capable of producing high resolution images. As the detection range increases multipath effects can become important. In this paper we present the results of at-sea tests in the Gulf of Mexico and St. Andrew Bay with a dual-frequency (20 kHz and 180 kHz) synthetic aperture sonar. Surprisingly, it was found that range to water depth ratios as small as 4 often produced multipath signals significantly stronger than direct path signals. These effects on the synthetic aperture processing and image quality are discussed.

Session 3pAO

Acoustical Oceanography: Acoustical Oceanography Prize Lecture

Peter F. Worcester, Chair
Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0225

Chair's Introduction—12:55

Invited Paper

1:00

3pAO1. Ocean tomography, inverse methods, and broadband ocean acoustics. Bruce D. Cornuelle (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093)

Ocean acoustic tomography, as proposed by Munk and Wunsch in 1979, and as implemented by the Ocean Tomography Group, uses ray travel times to estimate ocean sound speed and currents. Earlier work by Medwin (1970) and Hamilton (1977) used pulse travel times as measures of integrated sound speed along paths at short and long range, respectively. Munk and Wunsch (1979) recognized that broadband transmissions between many instruments could be used with inverse methods [Backus and Gilbert (1967); Liebelt (1967)] to reconstruct 3D ocean sound speed fields from the travel times along multiple paths. Inverse methods are widely used in ocean acoustics and in physical oceanography, and the modern challenge is to incorporate time dependence into the inverse methods to take advantage of the improving ocean dynamical models. In addition, the understanding of broadband acoustic propagation has improved to the point of refining the sensitivity kernel for travel time measurements beyond the simple geometrical optics of ray paths. This paper will review the evolving use of forward and inverse methods in acoustical oceanography, primarily with application to acoustic tomography.

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

Mardi C. Hastings, Chair
Biomedical Engineering Center, The Ohio State University, 1080 Carmack Road, Columbus, Ohio 43210

Chair's Introduction—2:00

Invited Papers

2:05

3pID1. Hot topics: Signal processing in acoustics. James Candy (Univ. of California, Lawrence Livermore Natl. Lab., P.O. Box 808, L-156, Livermore, CA 94551)

Signal processing represents a technology that provides the mechanism to extract the desired information from noisy acoustical measurement data. The desired result can range from extracting a single number like sound intensity level in the case of marine mammals to the seemingly impossible task of imaging the complex bottom in a hostile ocean environment. Some of the latest approaches to solving acoustical processing problems including sophisticated Bayesian processors in architectural acoustics, iterative flaw removal processing for non-destructive evaluation, time-reversal imaging for buried objects and time-reversal receivers in communications as well as some of the exciting breakthroughs using so-called blind processing techniques for deconvolution are discussed. Processors discussed range from the simple to the sophisticated as dictated by the particular application. It is shown how processing techniques are crucial to extracting the required information for success in the underlying application.

2:25

3pID2. Determining the sources of sounds: Psychological acoustics. William A. Yost (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626, wyost@luc.edu)

Perhaps the most important task performed by an organism's sensory system is determining objects in the world surrounding the organism. This is a challenging task for the auditory system, especially when many objects exist at the same time. The peripheral auditory system provides the neural code for the spectral-temporal structure of the complex sound impinging on the ears. It does not code for the sources of the sound. Neural centers beyond the periphery most analyze this spectral-temporal code in order to form perceptual images that make up the auditory scene. Recent psychoacoustical and perceptual research on sound source determination and segregation will be described. This research is just beginning to suggest ways in which the auditory scene is determined. [Work sponsored by NIDCD.]

2:45

3pID3. Hot topics in engineering acoustics. Stephen C. Thompson (Knowles Electron., LLC, 1151 Maplewood Dr., Itasca, IL 60143)

From the ASA Handbook, the technical committee on Engineering Acoustics is ''concerned with the evolution and improvement of acoustical techniques and apparatus, and with the promotion of new applications of acoustics for useful purposes.'' The interest of acousticians in apparatus and techniques is at least as old as the society, as can be verified by a quick scan of the first volume of the journal. Active development continues in areas that include new materials for high power sonar devices, new structures for hearing aid transducers, the announcement of the first commercial micromachined silicon microphone, multimedia and 3-D audio applications, and more. This paper is an abbreviated report on a few of these developments.

WEDNESDAY AFTERNOON, 5 JUNE 2002 KINGS GARDEN NORTH, 1:00 TO 2:00 P.M.

Session 3pSA

Structural Acoustics and Vibration: Holography

Sean F. Wu, Chair
Department of Mechanical Engineering, Wayne State University, 505 Anthony Wayne Drive, Detroit, Michigan 48202

Contributed Papers

1:00

3pSA1. Spherical wave approximation to arbitrary sound fields. Nassif E. Rayess (College of Eng. and Sci., The Univ. of Detroit, Mercy, Detroit, MI 48219-0900)

The ability to expand the acoustic field radiated from arbitrarily shaped objects in the form of a series of spherical wave functions has always been a point of interest. Despite the paucity of theoretical understanding, experimental and numerical work has repeatedly shown that such expansions can lead to relatively accurate approximations of the radiated sound fields. A general understanding of the problem as well as a quantitative approximation of the resulting errors are deduced from the sound radiation model based on a volume distribution of monopole sources. The errors involved in the approximation are a direct result of the presence of monopole sources outside the largest inscribed sphere. It is shown that under many circumstances, such errors are relatively small and can be mitigated further by the proper choice of the method used to determine the coefficients of the expansion. The collocation method with an appropriate sampling scheme has been used with a great deal of success to determine the expansion coefficients in acoustic holography problems.

1:15

3pSA2. HELS based near-field acoustic holography for a highly nonspherical structure. Manmohan Moondra and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

The Helmholtz Equation Least Squares (HELS) method [Wu and Yu, J. Acoust. Soc. Am. **104**, 2054–2060 (1998); Wu, *ibid.* **107**, 2511–2522 (2000)] are validated experimentally in both exterior and interior regions of a highly nonspherical structure. This is contrary to the common belief that expansion solutions based on spherical coordinates are valid within the region bounded by spheres. Detailed explanations for this phenomenon, known as the Rayleigh hypothesis, are given in a separate paper. Shown here are the experimental validations of the acoustic fields reconstructed by the HELS method for a cabinet with overall dimensions 2.05 m $\times$1.05 m $\times$0.75 m. A loudspeaker is placed inside the cabinet. The white noise produced by the speaker is measured by an array of 48 microphones stationed outside. The acoustic pressures in the near- and far-fields are reconstructed and validated with respect to those measured at the same locations. Also reconstructed is the normal component of the time-averaged acoustic intensity, from which the transmission paths are identified. Such transmission paths cannot be revealed based on the acoustic pressure distribution alone because pressure is a scalar quantity. The interior acoustic pressures are reconstructed with the speaker placed outside and microphone array inside the cabinet. [Work supported by NSF.]

1:30

3pSA3. Reconstruction of acoustic radiation from a finite object in nonfree space. Sean Wu and Xiang Zhao (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

The near-field acoustic holography (NAH) techniques developed so far have focused on reconstruction of acoustic radiation from a vibrating structure in either an unbounded exterior region or an enclosed region so that it can be solved as an interior problem. In practice, many vibrating structures are mounted on a solid foundation or in the vicinity of reflecting surfaces. For NAH to become a robust diagnostic tool, the effect of acoustic pressure reflection from nearby surfaces must be considered. In this paper, the Helmholtz equation least-squares (HELS) method [Wu, J. Acoust. Soc. Am. **107**, 2511–2522 (2000)] will be modified which expresses the field acoustic pressure as a superposition of both out-going and in-coming spherical waves. The latter will be used to approximate the acoustic pressure reflection from nearby surfaces. To guarantee a convergence of the reconstructed acoustic field, a constrained minimization will be conducted with respect to the expansion functions to determine the

optimal numbers of expansion terms for both out-going and in-coming waves. This concept will also be extended to reconstruction of acoustic radiation from an arbitrarily shaped vibrating structure for which the conventional HELS formulation has been shown to have difficulties to yield satisfactory reconstruction. [Work supported by NSF.]

1:45

3pSA4. On the behavior of solutions obtained by the HELS method inside the minimum sphere. Tatiana Semenova and Sean F. Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

Previous experimental and numerical studies have shown that the Helmholtz equation least-squares (HELS) method [Wu, J. Acoust. Soc. Am. **107**, 2511–2522 (2000)] can be an effective methodology for nearfield acoustical holography. However, results also demonstrated that the accuracy of reconstruction on a highly nonspherical surface might be unsatisfactory. The validity of the HELS method on these highly nonspherical surfaces has become a controversial topic, just like the Rayleigh hypothesis for acoustic scattering on a corrugated surface. In this paper, the validity of solutions obtained by the HELS method will be examined for acoustic radiation from an infinite column with a rectangular cross section. The acoustic field inside the minimal circle that encircles the column will be reconstructed. The convergence of solutions both on the surface and inside the minimum circle will be checked. Moreover, a modified HELS formulation will be used, which describes the acoustic pressure as a superposition of both out-going and in-coming spherical waves. The nature of the solutions thus obtained will be examined, and the dependence of convergence of the HELS solution on the validity of the Rayleigh hypothesis and on the measurement locations will be investigated. [Work supported by NSF.]

WEDNESDAY AFTERNOON, 5 JUNE 2002 GRAND BALLROOM 1, 10:00 TO 12:00 NOON

Plenary Session, Business Meeting, Awards Ceremony

William M. Hartmann, President
Acoustical Society of America

Business Meeting

Presentation of Certificates to New Fellows

Rachel K. Clifton
Kerry W. Commander
Brian H. Houston
Patrick M. Hurdle
Darlene R. Ketten
Birger Kollmeier
Asbjörn Krokstad
David E. Marsh
Andrzej Rakowski
David E. Marsh
Andrzej Rakowski
W. John Richardson
Nicholas Rott
Christopher Shera
Leon H. Sibul
Andrea M. Simmons
David C. Swanson
Lynne A. Werner

Announcement of Prize

Bruce Cornuelle, 2002 Medwin Prize in Acoustical Oceanography

Presentation of Awards

R. Bruce Lindsay Award to James M. Finneran

R. Bruce Lindsay Award to Thomas J. Royston

Silver Medal in Psychological and Physiological Acoustics to Neal F. Viemeister

Gold Medal to Robert E. Apfel

Gold Medal to Tony F. W. Embleton

Session 4aAA

Architectural Acoustics: Acoustics of Meeting Facilities

Jesse J. Ehnert, Chair
Newcomb & Boyd, 303 Peachtree Center Avenue, Northeast, Suite 525, Atlanta, Georgia 30303-1277

Chair's Introduction—8:00

Invited Papers

8:05

4aAA1. Convention facility acoustical design. David Marsh (Pelton Marsh Kinsella, 1420 W. Mockingbird Ln., Ste. 400, Dallas, TX 75247)

Convention centers usually include a variety of meeting spaces with widely varying acoustical requirements. Exhibit halls with several hundred thousand square feet of contiguous space are common along with divisible ballrooms and smaller meeting rooms. Some even have upscale boardrooms with built-in audio–video capabilities rivaling modern corporate facilities. This paper provides an overview of appropriate acoustical design criteria for typical occupied spaces in a convention center in terms of background noise levels, sound isolation between adjacent spaces, and acceptable reverberation times. Examples are given from real-world projects.

8:35

4aAA2. Engineering challenges of the acoustics of a political convention. Jack E. Randorff (Randorff and Assoc., 11 W. Canyon View Dr., Ransom Canyon, TX 79366)

The acoustical challenges encountered during the 2000 Republican Convention are discussed. The convention has held in Philadelphia, Pennsylvania's First Union Center. This venue is a dual-purpose facility catering to professional basketball and professional ice hockey. The acoustical needs of the delegates and the broadcast audience are discussed. The technical performance requirements of convention sound reinforcement and media network broadcast feed are outlined. The necessary technical and performance tradeoffs are enumerated with respect to the physical constraints, schedule requirements, budget limitations, and technical planning committee expectations. The conversion of a major sporting arena to a large-scale meeting room with reverberation times and general room conditions conducive to good listening was a significant undertaking. The site had been chosen for a preliminary screening visit approximately 2 years before. This presentation is a followup to ''Acoustics of Political Conventions—A Review,'' delivered at the Acoustical Society of America 139th Meeting in Atlanta in June 2000, 2 months before the convention in Philadelphia.

9:05

4aAA3. Optimal acoustical design of commercial meeting and teleconference rooms. Angelo Campanella (Campanella Assoc., 3201 Ridgewood Dr., Hilliard, OH 43026)

Along with a quiet NC-25 environment, speech communication among multiple individuals in a conference room, especially for teleconferencing, should be enhanced. Conferees are often located around a table and along the room perimeter. Speech propagation among all is enhanced by a flat ceiling reflector placed as low as practical over the table and the seated participants. This reinforces direct sound with a first-bounce reflection between all parties and microphones. Ceiling perimeter, 50%–75% of the ceiling area, is absorptive. This arrangement works for a speaker-phone and for microphones placed on the ceiling reflector avoiding paper shuffling noise. In the latter case, loudspeakers are placed on the absorbing perimeter acoustical tile. Absorption must be placed on the sidewalls to avoid flutter echoes. Uneven absorber distribution among surfaces is evaluated with the Fitzroy reverberation time [D. Fitzroy, J. Acoust. Soc. Am. **31**, 893–897 (1959)]. It was found that this RT at 500 Hz should be $[1+\log(V)]/10$ (about 1/2 s) to avoid a ''hollow'' timbre in purveyed speech. Examples of good and troublesome teleconference rooms are discussed.

9:35

4aAA4. The importance of early reflections for speech intelligibility in rooms. John Bradley and Hiroshi Sato (M27-IRC, Natl. Res. Council, Montreal Rd., Ottawa, ON K1A 0R6, Canada)

New results demonstrate that in rooms for unamplified speech communication early reflections are important for achieving adequate speech intelligibility and that early reflection energy is typically much greater in magnitude than the direct sound energy. Speech intelligibility was measured using a rhyme test in simulated sound fields that included a direct sound with varied early reflections and reverberant sound along with a constant level of ambient noise. The results confirmed that added early reflection energy is equivalent to increased direct sound energy. The combination of direct sound and early reflections increases the effective signal-to-noise ratio and the related speech intelligibility scores for both impaired and nonimpaired listeners. The new results also show, that for common conditions where the direct sound is reduced, it is only possible to understand speech because of the presence of early reflections. Analyses of measured impulse responses in rooms for speech, in terms of a measure of the benefit from early reflections, show that early reflections can increase the effective signal-to-noise ratio by up to 9 dB. The primary goal of room acoustics design should be to maximize early reflection energy to make possible increased effective signal-to-noise ratios.

Session 4aBB

Biomedical Ultrasound/Bioresponse to Vibration: Ultrasound Propagation in and Through Bone

Robin O. Cleveland, Chair
Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Chair's Introduction—8:00

Invited Papers

8:05

4aBB1. Experimental studies of the interaction of ultrasound with bone: What have we learned in 50 years? Patrick Nicholson (Depts. of Health Sci. and Phys., Univ. of Jyvaskyla, Finland)

Despite five decades of study, our knowledge of the interaction of ultrasound with bone remains incomplete. On the one hand, ultrasound propagation in cortical bone is relatively well-understood. Given its low porosity, cortical bone can be treated as a homogeneous anisotropic solid for the purposes of low-MHz ultrasound assessment. Consequently, ultrasound velocity measurements have been used very successfully as an *in vitro* tool to determine elastic constants. Developing useful clinical techniques for cortical bone assessment has been more difficult, although approaches such as critical angle reflectometry look promising. Understanding ultrasound propagation in trabecular bone, a highly porous fluid-saturated solid with a complicated microarchitecture, has proved a tougher problem. A wide range of clinical devices are already in use despite the fact that, with some notable exceptions, there have been few rigorous attempts to probe the underlying physics. This situation is being resolved, with recent studies into the validity of Biots and Schoenbergs theories for porous and stratified media, respectively, and also work looking at the role of scattering and the value of backscatter measurements. The insights emerging from this new phase of research are likely to trigger a new generation of acoustic techniques for bone characterization.

8:30

4aBB2. Ultrasonic scattering models for cancellous bone. Pascal Laugier, Frederic Padilla, and Frederic Jenson (Laboratoire d'Imagerie Parametrique, 15 rue de l'Ecole de Medecine, 75006 Paris, France)

In recent years, quantitative ultrasound (QUS) measurements have played a growing role in the assessment and management of osteoporosis. This development is attributable to the current wide availability of ultrasonic equipment measuring acoustic parameters in transmission at several skeletal sites. QUS provides equivalent fracture risk assessment compared to conventional x-ray absorptiometric techniques. It is generally accepted that transmission QUS represents a surrogate marker for bone mineral density. Several investigations are currently being conducted to assess innovative QUS techniques to determine and utilize the full potential of this technology for the benefit of detecting pathological conditions that affect bone strength. Among these developments, reflection and scattering techniques may be useful for the quantitative analysis of bone microarchitecture. We will review recent experimental data and theoretical modeling of ultrasound scattering by cancellous bone. Good agreement between experimental data and predictions using a weak scattering model has been reported by different authors for the magnitude and the frequency dependence of the backscatter coefficient. We also found good agreement between the experimental mean trabecular thickness derived from the 3D microarchitecture analysis and theoretical predictions. These results open interesting prospects for bone microarchitecture characterization from *in vivo* measurements.

8:55

4aBB3. Distinct US modes in trabecular bones: Properties and modeling. Mariusz Kaczmarek and Michal Pakula (Bydgoszcz Univ., Inst. of Env. Mech. & Appl. Comp. Sci., Chodkiewicza 30, 85-064 Bydgoszcz, Poland)

Marrow saturated trabecular bones are two-phase composites in which the propagation of ultrasonic waves is influenced by relative motion of phases. In particular, in studies *in vitro* two longitudinal wave modes are observed with specific behavior of attenuation and dispersion. Although the existence of the modes is predicted by the macroscopic model proposed by Biot some of their basic properties cannot be derived from the model [see Hosokawa and Otani (1997)]. Thus numerous modifications of the macroscopic model have been developed including layered structure [see Hughes *et al.* (1999)] or incorporating scattering [see Strelitzki *et al.* (1998)] in order to obtain better agreement with experimental results. Still, however, some important features such as high attenuation of the fast wave, insignificant dispersion and moderate attenuation of the slow wave, for relatively short waves as compared with the characteristic size of pores or trabeculae are not well described. This paper discusses predictions of the proposed microscopic (or exact) two-phase model developed for chosen cellular geometries [see Kaczmarek *et al.* (2002)]. The comparison of theoretical results with our own and other experimental data is performed. Additionally, references to results obtained for model material (phantoms) with structure corresponding to the models adopted are made.

9:20

4aBB4. Dependence of ultrasonic scattering on frequency and microarchitecture in trabecular bone: Theory and experiment. Keith A. Wear (Food and Drug Administration, 12720 Twinbrook Pkwy., Rockville, MD 20857)

Measurements of ultrasonic properties of calcaneus (heel bone) have been shown to be effective for the diagnosis of osteoporosis. However, the mechanisms underlying the interaction between ultrasound and bone are currently not well understood. A model that predicts backscatter from trabecular bone has been developed. Scattering is assumed to originate from the surfaces of trabeculae, which are modeled as long, thin, elastic cylinders with radii small compared with the ultrasonic wavelength. Experimental measurements of backscatter using broadband ultrasound centered at 500 kHz from 43 trabecular bone samples (from human calcaneus) *in vitro* have been performed. Microcomputed tomography has been performed on all 43 samples in order to measure microarchitectural features. The theory correctly predicts the measured dependences of backscatter on ultrasonic frequency and trabecular thickness. [Funding from the FDA Office of Womens Health is gratefully acknowledged.]

9:45

4aBB5. Continuous mapping of the acoustic properties of biological materials with emphasis on bone. Sidney Lees (The Forsyth Inst., Boston, MA 02115, slees@forsyth.org)

Instruments determine the time of flight through a prepared section of bone of uniform thickness. A multigated receiver can be set to capture successive pulse echoes. A clock turned on by the first pulse echo is turned off by the next. A faster driven transducer maps the bone section. Several tissue properties can be calculated from the time of flight, the amplitude, thickness of the bone section and its density. In particular, the variation of the elastic properties provides a detailed picture of stress distribution on the bone. Several examples will be shown.

10:10–10:25 Break

10:25

4aBB6. QUS devices for assessment of osteoporosis. Christian Langton (Ctr. for Metabolic Bone Disease, Univ. of Hull and Hull & East Yorkshire Hospitals Trust, Hull HU3 2RW, UK)

The acronym QUS (Quantitative Ultrasound) is now widely used to describe ultrasound assessment of osteoporosis, a disease primarily manifested by fragility fractures of the wrist and hip along with shortening of the spine. There is currently available a plethora of commercial QUS devices, measuring various anatomic sites including the heel, finger, and tibia. Largely through commercial rather than scientific drivers, the parameters reported often differ significantly from the two fundamental parameters of velocity and attenuation. Attenuation at the heel is generally reported as BUA (broadband ultrasound attenuation, the linearly regressed increase in attenuation between 200 and 600 kHz). Velocity derivatives include bone, heel, TOF, and AdV. Further, velocity and BUA parameters may be mathematically combined to provide proprietary parameters including ''stiffness'' and ''QUI.'' In terms of clinical utility, the situation is further complicated by ultrasound being inherently dependent upon ''bone quality'' (e.g., structure) in addition to ''bone quantity'' (generally expressed as BMD, bone mineral density). Hence the BMD derived WHO criteria for osteoporosis and osteopenia may not be directly applied to QUS. There is therefore an urgent need to understand the fundamental dependence of QUS parameters, to perform calibration and cross-correlation studies of QUS devices, and to define its clinical utility.

Contributed Papers

10:50

4aBB7. On the relationship of ultrasonic properties to density and architecture in trabecular bone. Patrick Nicholson and Mary Bouxsein (Orthopedic Biomechanics Lab., BIDMC and Harvard Med. School, Boston, MA 02115)

As previously reported elsewhere, we have made ultrasonic measurements in human trabecular bone and have explored relationships with microstructural properties, the latter derived from microcomputed tomography. However, multicollinearity in these data means that conventional regression analysis cannot reliably identify the underlying causal relationships. In an effort to move beyond such limitations, we used our experimental data to test some models of possible interactions between ultrasound and bone. In particular, we compared several models for predicting acoustic velocity in two-phase media as a function of the bone volume fraction. We found good agreement only with the theory of Kuster and Toksoz [Geophysics **39**, 587 (1974)] based on scattering by an effective medium. Turning our attention to attenuation, we examined relationships with trabecular thickness (Tb.Th) and the number of trabeculae per unit volume (Tb.N). The exponent relating attenuation to Tb.Th was 3.2, lower than the value of 4 predicted for long wavelength inelastic scattering by cylinders. This may be due to multiple scattering, since restricting the analysis to specimens with relatively low Tb.N yielded an exponent of 3.9. The exponent relating attenuation to Tb.N was 1.3, higher than the expected value of unity, which may again reflect the influence of multiple scattering.

11:05

4aBB8. Development and validation of a multiecho computer simulation of ultrasound propagation through cancellous bone. Christian Langton and Luke Church (Ctr. for Metabolic Bone Disease, Univ. of Hull and Hull & East Yorkshire Hospitals Trust, Hull HU3 2RW, UK)

Cancellous bone consists of a porous open-celled framework of trabeculae interspersed with marrow. Although the measurement of broadband ultrasound attenuation (BUA) has been shown to be sensitive to osteoporotic changes, the exact dependence on material and structural parameters has not been elucidated. A 3-D computer simulation of ultrasound propagation through cancellous bone has been developed, based upon simple reflective behavior at the multitude of trabecular/marrow interfaces. A cancellous bone framework is initially described by an array of bone and marrow elements. An ultrasound pulse is launched along each row of the model with partial reflection occurring at each bone/marrow interface. If a reverse direction wave hits an interface, a further forward (echo) wave is created, with phase inversion implemented if appropriate. This process is monitored for each wave within each row. The effective received signal is created by summing the time domain data, thus simulating detection by a phase-sensitive ultrasound transducer, as incorporated in clinical systems. The simulation has been validated on a hexagonal honeycomb design of variable mesh size, first against a commercial computer simulation solution (Wave 2000 Pro), and second, via experimental measurement of physical replicas produced by stereolithography.

11:20

4aBB9. Acoustic properties of bovine cancellous bone from 0.5 to 1.5 MHz.[a] Kang Il Lee, Heui-Seol Roh, and S. W. Yoon (Acoust. Res. Lab., Dept. of Phys., Sung Kyun Kwan Univ., Suwon 440-746, Republic of Korea, swyoon@skku.ac.kr)

Most previous studies using ultrasound for the diagnosis of osteoporosis have employed ultrasound in the frequency range from 0.2 to 1 MHz. Acoustic properties of 14 defatted bovine cancellous bone specimens were investigated *in vitro*. Speed of sound (SOS) and broadband ultrasonic attenuation (BUA) were measured using two matched pairs of transducers with the center frequencies of 1 and 2.25 MHz, respectively, in order to cover a broad frequency range from 0.5 to 1.5 MHz. In this frequency range, SOS and BUA show significant linear positive correlations with apparent bone density. These results suggest that the frequency range from 0.5 to 1.5 MHz may also be useful in the diagnosis of osteoporosis. [Work supported by BK21 Program, KRF (KRF-2000-015-DP0718), and KOSEF (KOSEF-2000-0238-100) in Korea.] [a] For (Biomedical Ultrasound/Bioresponse to Vibration) Best Student Paper Award.

11:35

4aBB10. A noninvasive method for focused ultrasound surgery through the human skull. Gregory Clement and Kullervo Hynynen (Dept. of Radiol., Harvard Med. School and Brigham and Women's Hospital, Boston, MA 02115)

A technique for focusing ultrasound through the skull bone is described and verified. The approach is based on a layered wave-vector-frequency domain model, which simulates ultrasound propagation through the skull bone using input from CT scans of the head. The algorithm calculates the driving phase of each element in a transducer array in order to maximize the signal at the intended focus. This approach is tested on ten *ex vivo* human skulls using a 500-element hemispherical array operated at 0.74 MHz. A stereotaxic reference frame is affixed to the skulls in order to provide accurate registration between the CT images and the transducer. The focal quality is assessed with a hydrophone placed inside of the skull. In each trial the phase correction algorithm successfully restored the focus inside the skull in a location within 1 mm from the intended focal point. Focusing at high powers ($>$800-W electrical input) is demonstrated using a brain phantom placed inside a skull. The results demonstrate the feasibility of using the method for completely noninvasive ultrasound brain surgery and therapy.

THURSDAY MORNING, 6 JUNE 2002 BRIGADE ROOM, 9:00 TO 11:50 A.M.

Session 4aEA

Engineering Acoustics and Underwater Acoustics: Mine Hunting Sonar

Thomas R. Howarth, Chair
NAVSEA Newport, 1176 Howell Street, Newport, Rhode Island 02841

Chair's Introduction—9:00

Invited Papers

9:05

4aEA1. Minehunting sonar system research and development. Brian Ferguson (Maritime Operations Division–Sydney, Defence Sci. and Technol. Organisation, P.O. Box 44, Pyrmont NSW 2009, Australia)

Sea mines have the potential to threaten the freedom of the seas by disrupting maritime trade and restricting the freedom of maneuver of navies. The acoustic detection, localization, and classification of sea mines involves a sequence of operations starting with the transmission of a sonar pulse and ending with an operator interpreting the information on a sonar display. A recent improvement to the process stems from the application of neural networks to the computed aided detection of sea mines. The advent of ultrawideband sonar transducers together with pulse compression techniques offers a thousandfold increase in the bandwidth-time product of conventional minehunting sonar transmissions enabling stealth mines to be detected at longer ranges. These wideband signals also enable mines to be imaged at safe standoff distances by applying tomographic image reconstruction techniques. The coupling of wideband transducer technology with synthetic aperture processing enhances the resolution of side scan sonars in both the cross-track and along-track directions. The principles on which conventional and advanced minehunting sonars are based are reviewed and the results of applying novel sonar signal processing algorithms to high-frequency sonar data collected in Australian waters are presented.

9:35

4aEA2. Current development of high-resolution MCM minehunting sonars for small AUVs. Kerry W. Commander, James T. Christoff, and Joseph L. Lopes (Coastal Systems Station, Code R21, Panama City, FL 32407-7001, commanderkw@ncsc.navy.mil)

The U.S. Navy recently awarded contracts to develop a new generation of autonomous underwater vehicles (AUVs) and high-performance acoustic sensors that are compatible with these vehicles. These new AUV systems will place minehunting crews and marine mammals well out of dangerous waters in the future. The requirement that these vehicles be small, durable, and easily launched and recovered places severe constraints on the sensor system payloads. The sensor systems must be rugged, low-powered, and small, yet effective enough to justify the risk to the vehicle. They must also have very low false alarm rates (high clutter rejection) because reacquisition and neutralization are costly, time consuming, and closely related to the number of targets found. These new

acoustic sensor systems will use new broadband techniques along with high-resolution multi-aspect designs to achieve their goals from such small platforms. The acoustic sensor packages will consist of advanced very high-resolution Ahead Looking Sonars (ALS) and very high-resolution Synthetic Aperture sonars (SAS). This paper will review the new sonar system designs being considered by the Navy.

10:05

4aEA3. Multi-platform sonar concepts for buried target detection and classification. Joseph R. Edwards, Te-Chih Liu, and Henrik Schmidt (Dept. of Ocean Eng., MIT, 77 Massachusetts Ave., Rm. 5-435, Cambridge, MA 02139)

Buried target classification is of paramount importance in mine countermeasures (MCM) applications. The Generic Oceanographic Array Technology Sonar (GOATS) project approaches this fundamental problem by exploiting the assets of autonomous underwater vehicles (AUVs). Recent developments in unmanned vehicle technology have opened the door for new sonar design concepts that may be applied to create alternative target detection and classification methods in complex environments. Adaptive vehicle behaviors provide the potential for the sonar system to adjust its geometry for optimal performance. The vehicles can also access very shallow water environments, which allows for more complete multi-aspect views of targets of interest than are possible with remote towed array systems. In this paper, it is demonstrated that the multi-platform system can provide improved detection and classification capability, while at the same time reducing the computational requirements. Data from prior GOATS experiments (1998 and 2000) are used to illustrate the effectiveness of the multi-platform sonar concepts, and real-time simulations are used to show the planned implementation of these techniques on-board the AUVs for GOATS'02.

10:35–10:50 Break

10:50

4aEA4. Toward a robust target scattering measurement sonar for low frequencies. Brian H. Houston, H. J. Simpson, S. Liskey, T. Yoder, and J. A. Bucaro (Naval Res. Lab., 4555 Overlook Ave., Washington, DC 20375)

In order to develop sonar systems that are non-image-based and that take advantage of target structural acoustics features, broadband capabilities have to be created to operate in shallow water and in the nontraditional frequency band below 30 kHz. Here, source wavelength–aperture ratios are high, resulting in significant source divergence and multipath. At the lowest frequencies, evanescent wave penetration is important in the detection of structures buried below the sediment–fluid interface. Further, the recovery of scattered energy requires nontraditional vertical apertures and synthetic array techniques that have both similarities and differences when compared to high-frequency imaging-based systems. The differences include relatively large resolution cell sizes, robustness to vehicle motion, and the clear potential for higher coverage rates. Towards this end, we discuss the precise measurement of low-frequency scattering cross section in shallow water using an UUV-based synthetic array approach. Broadband environmental acoustic propagation data taken in a bay environment will be presented, together with synthetic aperture data taken with a 20-m horizontal rail. The results of these measurements and the implications for UUV-based broadband sonars will be discussed.

11:20

4aEA5. Obstacle avoidance sonar for submarines. Albert C. Dugas and Kenneth M. Webman (Naval Undersea Warfare Ctr. Div. Newport, 1176 Howell St., Newport, RI 02841-1708, mankm@npt.nuwc.navy.mil)

The Advanced Mine Detection Sonar (AMDS) system was designed to operate in poor environments with high biological and/or shallow-water boundary conditions. It provides increased capability for active detection of volume, close-tethered, and bottom mines, as well as submarine and surface target active/passive detection for ASW and collision avoidance. It also provides bottom topography mapping capability for precise submarine navigation in uncharted littoral waters. It accomplishes this by using advanced processing techniques with extremely narrow beamwidths. The receive array consists of 36 modules arranged in a 15-ft-diameter semicircle at the bottom of the submarine sonar dome to form a chin-mounted array. Each module consists of 40 piezoelectric rubber elements. The modules provide the necessary signal conditioning to the element data prior to signal transmission (uplink) through the hull. The elements are amplified, filtered, converted to digital signals by an A/D converter, and multiplexed prior to uplink to the inboard receiver. Each module also has a downlink over which it receives synchronization and mode/gain control. Uplink and downlink transmission is done using fiberoptic telemetry. AMDS was installed on the USS Asheville. The high-frequency chin array for Virginia class submarines is based on the Asheville design.

Session 4aMU

Musical Acoustics and Psychological and Physiological Acoustics: Music Recognition Systems I

James W. Beauchamp, Chair
School of Music, Department of ECE, University of Illinois at Urbana-Champaign, 1114 West Nevada, Urbana, Illinois 61801

Chair's Introduction—8:30

Invited Papers

8:35

4aMU1. The existence region for melodic pitch and computational models. Roy D. Patterson, Katrin Krumbholz (Ctr. for the Neural Basis of Hearing, Dept. of Physiol., Univ. of Cambridge, Downing St., Cambridge CB2 3EG, UK), and Daniel Pressnitzer (Ircam-CNRS, 75004 Paris, France)

The pitch of a note is closely related to the repetition rate of the acoustic wave, or the spacing between the harmonics in the Fourier spectrum of the wave. However, algorithms that extract the repetition rate or harmonic spacing from the sound itself, without regard for auditory processing, would assign pitch perceptions to notes that musicians would not attempt to use to convey melody. If a transcription algorithm is to correctly reject notes that humans would not use, it must include knowledge of the complex constraints that govern the existence region for melodic pitch. Similarly, composers designing synthetic tonal instruments need to know whether the listener will hear the melody the instrument is intended to convey. Psychophysical studies have recently been used to delimit the region of melodic pitch within the overall domain of pitch perception, and to distinguish melodic pitch from rattle pitch, rumble pitch, and flutter pitch, none of which conveys melody successfully. This paper describes how time-domain computational models of auditory processing explain the existence region of melodic pitch, and how they might be used to improve melody transcription and instrument design. [Work supported by the UK Medical Research Council, G9901257.]

9:00

4aMU2. Fundamental frequency estimation of singing voice. Alain de Cheveigné (CNRS, Ircam, 1 Pl. Igor Stravinsky, 75004 Paris, France, cheveign@ircam.fr) and Nathalie Henrich (CNRS, LAM, 11 Rue de Lourmel, 75015 Paris, France)

A method of fundamental frequency ($F0$) estimation recently developped for speech [de Cheveigné and Kawahara, J. Acoust. Soc. Am. (to be published)] was applied to singing voice. An electroglottograph signal recorded together with the microphone provided a reference by which estimates could be validated. Using standard parameter settings as for speech, error rates were low despite the wide range of $F0$s (about 100 to 1600 Hz). Most ''errors'' were due to irregular vibration of the vocal folds, a sharp formant resonance that reduced the waveform to a single harmonic, or fast $F0$ changes such as in high-amplitude vibrato. Our database (18 singers from baritone to soprano) included examples of diphonic singing for which melody is carried by variations of the frequency of a narrow formant rather than $F0$. Varying a parameter (ratio of inharmonic to total power) the algorithm could be tuned to follow either frequency. Although the method has not been formally tested on a wide range of instruments, it seems appropriate for musical applications because it is accurate, accepts a wide range of $F0$s, and can be implemented with low latency for interactive applications. [Work supported by the Cognitique programme of the French Ministry of Research and Technology.]

9:25

4aMU3. Separation of musical instruments based on amplitude and frequency comodulation. Barry D. Jacobson (Harvard-MIT Div. of Health Sci. and Technol., Cambridge, MA), Gert Cauwenberghs (Dept. of Elec. and Computer Eng., Johns Hopkins Univ.), and Thomas F. Quatieri (MIT Lincoln Lab.)

In previous work, amplitude comodulation was investigated as a basis for monaural source separation. Amplitude comodulation refers to similarities in amplitude envelopes of individual spectral components emitted by particular types of sources. In many types of musical instruments, amplitudes of all resonant modes rise/fall, and start/stop together during the course of normal playing. We found that under certain well-defined conditions, a mixture of constant frequency, amplitude comodulated sources can unambiguously be decomposed into its constituents on the basis of these similarities. In this work, system performance was improved by relaxing the constant frequency requirement. String instruments, for example, which are normally played with vibrato, are both amplitude and frequency comodulated sources, and could not be properly tracked under the constant frequency assumption upon which our original algorithm was based. Frequency comodulation refers to similarities in frequency variations of individual harmonics emitted by these types of sources. The analytical difficulty is in defining a representation of the source which properly tracks frequency varying components. A simple, fixed filter bank can only track an individual spectral component for the duration in which it is within the passband of one of the filters. Alternatives are therefore explored which are amenable to real-time implementation.

9:50

4aMU4. Separation of harmonic sounds using multipitch analysis and linear models for the overtone series. Tuomas Virtanen and Anssi Klapuri (Signal Processing Lab., Tampere Univ. of Technol., P.O. Box 533, FIN-33101 Tampere, Finland, tuomasv@cs.tut.fi)

A signal processing method for the separation of concurrent harmonic sounds is described. The method is based on a two-stage approach. First, a multiple fundamental frequency estimator is applied to find initial sound parameters which are reliable, but inaccurate and static. Second, time-varying sinusoidal parameters are estimated in an iterative algorithm. The harmonic structure is retained by keeping the frequency ratio of overtones constant over time. Overlapping harmonic components are resolved using linear models for the overtone series. In practice, the models retain the spectral envelope continuity of natural sounds. Simulation experiments were carried out using generated test signals, which were random mixtures of two to six notes from recorded natural instruments. The system is able to produce meaningful results in all polyphonies, the quality of separated sounds gradually degrading along with the polyphony. Some denoising algorithms were applied to suppress nonstationary noise component, such as drums in real-world music signals. However, the usability of the system for real musical signals is still quite limited.

10:15–10:25 Break

10:25

4aMU5. Comparison of machine and human recognition of isolated instrument tones. Ichiro Fujinaga (Peabody Conservatory of Music, Johns Hopkins Univ., Baltimore, MD 21202)

This paper describes three different machine recognition experiments and a recently conducted human experiment in order to compare the abilities of machines and humans to recognize isolated instrument tones. The computer recognition software is based on the Lazy Learning Machine, which is an exemplar-based learning system using a k-nearest neighbor (k-NN) classifier with a genetic algorithm to find the optimal set of weights for the features to improve its performance. The performance of the software was progressively improved by adding more features. These include centroid and other higher order moments, such as skewness and kurtosis, the velocities of moments, spectral irregularity, tristimulus, and time–domain envelope shape. Also, realtime recognition is now possible by using Miller Pucketts PD, a realtime software synthesis system, and his fiddle~ object. The training data was taken from the McGill Master Samples. The human experiment involved eighty-eight conservatory students. Although the average human scores are similar to the machine scores, the best human subjects far exceeded the capabilities of the machine. The excellent performance of the humans in this experiment presents new challenges for timbre-recognition computer models.

10:50

4aMU6. Offline and online tempo detection and beat tracking. Jean Laroche (Creative ATC, 1500 Green Hills Rd., Scotts Valley, CA 95066)

Tempo detection and beat tracking consist of measuring the tempo and estimating beat locations in musical signals. Online beat tracking refers to situations where the tempo at a given instant is estimated with access to a limited portion of the future at this time. In offline beat tracking, the whole track is available for processing, which makes the task somewhat easier. Our algorithms use a short-time Fourier analysis of the signal to extract a time-varying measure of the energy variations in several frequency subbands, resulting in an energy flux signal showing large values around note onsets, percussion hits, and other musically relevant events. At regular intervals (every second), the flux signal is correlated with a series of periodic discrete pulses corresponding to various candidate tempos and beat locations. The tempo- and beat-location candidates showing the best fit are retained for the next stage, where a dynamic programming algorithm is used to select the best series of beat locations, based on a measure of tempo continuity, and consistency between tempo and beat. Dynamic programming is used both for offline and online processing with fairly good results.

11:15

4aMU7. Robust identification/fingerprinting of audio signals using spectral flatness features. Juergen Herre, Eric Allamanche, Oliver Hellmuth, and Thorsten Kastner (Fraunhofer IIS-A. Am Weichselgarten 3, D-91058 Erlangen, Germany, hrr@iis.fhg.de)

Recently, the problem of content-based identification material has received increased attention as an important technique for managing the ever-increasing amount of multimedia assets available to users today. This talk discusses the problem of robust identification of audio signals by comparing them to a known reference (''fingerprint'') in the feature domain. Desirable properties of the underlying features include robustness with respect to common signal distortions and compactness of representation. A family of suitable features with favorable properties is described and evaluated for their recognition performance. Some applications of signal identification are discussed, including MPEG-7 Audio.

Session 4aPAa

Physical Acoustics: Thermoacoustics

Ray Scott Wakeland, Chair
Graduate Program in Acoustics, Pennsylvania State University, Applied Science Building, University Park, Pennsylvania 16804

Contributed Papers

8:00

4aPAa1. Efficiency of acoustic power sources from 200 to 20,000 watts. Philip S. Spoor and John A. Corey (CFIC, Inc., 302 10th St., Troy, NY 12180)

The emergence of acoustic ''pulse-tube'' refrigerators with no moving parts, and with the capacity to provide tens, hundreds, and even thousands of watts of cooling at cryogenic temperatures, is generating considerable commercial interest. To become economically feasible, these cryocoolers require efficient, compact, and inexpensive sources of acoustic power. The authors will discuss a series of five acoustic drivers available from CFIC that range in power from 200 watts to 20,000 watts, using opposing pairs of STAR™ linear motors. Efficiency data for units throughout the range will be presented, showing the relative importance of different loss mechanisms for each size. In particular we will focus on I^2R losses, which are well understood in the small motors but seem anomalously high in the larger motors.

8:15

4aPAa2. Absence of excess attenuation in Celcor ceramic at low frequencies. Jin Liu, Steven L. Garrett (Grad. Prog. in Acoust., Penn State Univ., University Park, PA 16804, jyl118@psu.edu), Gregory S. Long and Ayusman Sen (Penn State Univ., University Park, PA 16802)

Several studies of the excesses attenuation due to the porosity of Celcor™ material have been published [J. M. Sabatier *et al.,* J. Acoust. Soc. Am. **99**, 2430–2432 (1996)]. Since we were planning to use Celcor™ in a refrigeration application, we were concerned with this reported excess loss mechanism. To provide an independent measurement of the loss at our frequencies of interest, we placed a 38 mm long sample of Celcor™ at the center of a 700 mm long electrodynamically-driven plane wave resonator. Using the even modes, as per the technique described by Moldover *et al.* [*Modern Acoustical Techniques for the Measurement of Mechanical Properties* (Academic, New York, 2001), pp. 395–399], the quality factors were measured between 230 Hz and 1.6 kHz in air at atmospheric pressure on a pristine sample. Pores (but not channels) of that sample were then blocked by a sealing process that involved an initiator, azobisisobutronitrile (AIBN), mixed with the monomer hydroxyethyl methacylate. There was no observed decrease in attenuation of even modes for the coated sample below 1.0 kHz. The volume exclusion due to pore sealing was inferred by the increase in even mode resonance frequencies. Thermoviscous losses were computed with DELTAE and compared to measured losses. [Work supported in part by ONR and the United Technologies Corporation Professorship Endowment.]

8:30

4aPAa3. Wet wall thermoacoustics in a single pore. Brian Espino and Larry Wilen (Dept. of Phys. and Astron., Ohio Univ., Athens, OH 45701)

We will discuss thermoacoustic measurements in a single circular pore with a mixture of air and methanol vapor as the working gas. The vapor is in coexistence with liquid methanol capillary-condensed in a thin porous material lining the inside wall of the pore. From measurements with no temperature gradient, the diffuso-acoustic function $F(\lambda_D)$ may be obtained. At low frequencies, condensation/evaporation and vapor diffusion is rapid enough to maintain the methanol partial pressure constant at its saturated equilibrium value. At high frequencies, the methanol partial pressure undergoes adiabatic pressure oscillations along with the air. $F(\lambda_D)$ describes the transition between these behaviors. $\lambda_D = \sqrt{2}R/\delta_D$, where δ_D is the penetration depth for binary diffusion. The measurements probe a range of λ_D from about 1 to 10. With an applied temperature gradient, thermoacoustic effects due to condensation/evaporation combined with vapor transport are detected. Results will be compared to recent theoretical predictions for wet wall thermoacoustic effects [Slaton *et al.*, J. Acoust. Soc. Am. **105**, 65–73 (1999) and references cited within]. [Work supported by ONR.]

8:45

4aPAa4. Measurement of Rayleigh streaming in high-amplitude standing waves using laser Doppler anemometry. Michael W. Thompson and Anthony A. Atchley (Grad. Prog. in Acoust., Penn State Univ., 217 Applied Science Bldg., University Park, PA 16802, mwt-acs@psu.edu)

Current theories for Rayleigh streaming in a long, cylindrical pipe predict dc velocities which increase as the square of the amplitude of the first harmonic, and exhibit a parabolic radial profile. The velocity field in such a pipe, driven at one end and rigidly capped at the other, has been measured using LDA (laser Doppler anemometry) techniques. Analysis of the data produced by the LDA system can be complicated by two things: (i) the data are randomly sampled at a frequency below the Nyquist rate, resulting in a large, uneven sample spacing which is unsuitable for standard discrete Fourier transform techniques; and (ii) the dc velocity is typically two to three orders of magnitude smaller than the ac velocity, resulting in a very low signal-to-noise ratio. A robust signal processing algorithm is presented which overcomes these two obstacles, permitting fast and accurate estimates of the magnitudes, phases, and uncertainties of the velocity Fourier coefficients. The measured dc velocities agree well with the theoretical predictions when the tube is driven at low amplitudes, but deviate significantly from the theoretical predictions when the tube is driven at high amplitudes. [Work supported by ONR.]

9:00

4aPAa5. Effect of thermal conductivity on acoustic streaming in a channel. Mark F. Hamilton, Yurii A. Ilinskii, and Evgenia A. Zabolotskaya (Dept. of Mech. Eng., Univ. of Texas, Austin, TX 78712-1063)

Acoustic streaming generated in a viscous fluid with thermal conductivity is considered analytically and numerically. The fluid is confined to a channel formed by two parallel plates. The streaming pattern is similar to that in a viscous fluid without thermal conductivity. Specifically, an acoustic standing wave generates inner and outer streaming vortices that circulate in opposite directions. The outer vortex resembles Rayleigh streaming in a viscous fluid without thermal conductivity, and the inner vortex, confined mostly to the boundary layer region, is due to viscous stresses near the wall. Thermal conductivity is found to increase the streaming velocity in a gas by as much as 20%. Dependence of the viscosity coefficient on temperature is also taken into account, and this effect is found to reduce the streaming velocity. However, the first effect dominates the second, and therefore the net result is larger streaming velocities than in the absence of both effects. An analytical expression for acoustic streaming in both nar-

row and wide channels was derived and used to investigate the two effects individually. The analytical solution is in close agreement with numerical results based on the fully nonlinear equations of motion. [Work supported by ONR.]

9:15

4aPAa6. Analytical solution of the regenerator temperature distribution in a thermoacoustic device with mean flow. Nathan T. Weiland and Ben T. Zinn (Schools of Aerosp. and Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, gte852f@prism.gatech.edu)

Previous work by the authors has identified a potential for improvement in the efficiency of a traveling wave thermoacoustic engine by replacing the hot heat exchanger with a steady flow of hot gas. An essential step in determining the feasibility of such an engine lies in determining how the mean flow affects the temperature distribution in the regenerator, which in turn affects its acoustic driving capability. In the current work, the second order time-averaged thermoacoustic energy equation with mean flow is solved analytically for the mean temperature distribution in the regenerator. The resulting expression accounts for the dependence of the working fluid's thermal conductivity and viscosity on temperature, and results in a nearly exponential temperature profile. The specific effects of mean flow, axial conduction, acoustic velocity amplitude, and gas property variation on the temperature distribution in the regenerator will be explored. In turn, the effect of the temperature profile on regenerator performance will also be reviewed. In addition to their usefulness in the feasibility study, these results can also be applied to assess losses in existing thermoacoustic engines and refrigerators where acoustic streaming results in unwanted net mass flux through the regenerator.

9:30

4aPAa7. Minor-loss coefficients for the exit from heat exchangers in thermoacoustics. Ray Scott Wakeland and Robert M. Keolian (Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804, wakeland@psu.edu)

Heat exchanger minor losses in thermoacoustics are sometimes estimated using $\Delta\dot{E} = -(4/3\pi)(\rho_m A/2)K|u_A|^3$, where ρ_m, u_A, and A are the gas density, peak velocity amplitude, and cross-sectional area, and K is a ''minor loss coefficient'' [G. W. Swift and D. C. Ward, J. Thermophys. Heat Transfer **10**, 652–662 (1996)]. Because values for K are generally unavailable for oscillating flow, steady-flow results are used, including the formula $K_e = (1 - A_1/A_2)^2$ for sudden increase from cross-sectional area A_1 to a larger area A_2, such as at the exit from a heat exchanger. This steady-flow result assumes a *uniform* cross-sectional velocity profile. In this presentation, non-uniform, time-dependent velocity profiles for laminar oscillating flow are employed to calculate effective minor-loss exit coefficients, which might represent an improvement over the steady-flow formula for calculating flow losses at the exit from heat exchangers in thermoacoustic devices. Differences from the uniform flow calculation can be significant. For example, for $A_1/A_2 = 0.66$, a reasonable value for a heat exchanger, this calculation of minor loss is as much as 3.4 times greater than the uniform profile result. [Work supported by ONR and the Pennsylvania Space Grant Consortium.]

9:45

4aPAa8. Minor losses and streaming in thermoacoustic devices. Said Boluriaan and Philip J. Morris (Dept. of Aerosp. Eng., Penn State Univ., 233 Hammond Bldg., University Park, PA 16802)

The efficiency of thermoacoustic devices is limited by streaming and viscous losses. Viscous losses are due to energy dissipation in the vicinity of solid boundaries and minor losses. Minor losses are an additional loss of energy associated with transitions between channels and changes in flow direction. The physics behind minor losses in the high-amplitude oscillatory flows encountered in thermoacoustic devices is not well understood and estimates of its value are often based on results for steady flow. Furthermore, the presence of streaming adds to the complexity in the assessment of minor losses. In the current research, a sudden change in cross-sectional area of a resonator is used as a model to study streaming and minor losses. A time-accurate numerical simulation based on the Navier–Stokes equations is performed to calculate the flow variables in an oscillatory flow inside a resonator with a sudden expansion/contraction. A correlation between the pressure loss due to minor losses and different flow parameters involved in the problem is presented. Comparisons are also made with estimates of minor losses based on a quasisteady flow assumption. [Research supported by ONR.]

THURSDAY MORNING, 6 JUNE 2002 STERLINGS ROOMS 2 AND 3, 10:30 A.M. TO 12:00 NOON

Session 4aPAb

Physical Acoustics: Acoustic-Seismic Coupling

Howard B. Wheeler, Cochair
National Center for Physical Acoustics, University of Mississippi, Coliseum Drive, University, Mississippi 38677

James M. Sabatier, Cochair
National Center for Physical Acoustics, University of Mississippi, Coliseum Drive, University, Mississippi 38677

Contributed Papers

10:30

4aPAb1. Resonant frequency response and surface particle velocity profiles for buried land mines: Theory and experiment. James M. Sabatier (U.S. Army CECOM Night Vision and Electron. Sensors Directorate, Ft. Belvoir, VA 22060), Doru Velea (Planning Systems, Inc., Reston, VA 20191), Ning Xiang, and Roger Waxler (Univ. of Mississippi, University, MS 38677)

Acoustic-to-seismic coupling is currently used to detect buried landmines by measuring a contrast in the particle velocity of the air–soil interface directly above the mine (on-target) and away from the mine (off-target). Field measurements reveal a resonance phenomenon of the mine–soil system at frequencies around 100 Hz. The resonance frequency and the spatial profile of the on/off-target velocity ratio depend on the type of the mine, depth, and soil characteristics. Experimental results for antitank mines will be presented. These results will be compared with predictions by a recently developed model for the scattering of normally incident sound off of a landmine. The model assumes that the mine has a compliant top and the soil is an effective fluid. [Work supported by the U.S. Army Communications-Electronics Command Night Vision and Electronic Sensors Directorate and the U. S. Army Research Office.]

10:45

4aPAb2. Analysis of surface wave formation over a grating surface with bounded beams. Wenhao Zhu (Inst. of Microstructural Sci., Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

It had been demonstrated previously that homogeneous plane waves could not excite effective surface wave over periodic surface structures at low frequencies. In this work, acoustic surface wave over a periodic surface is predicted under limited source incidence (e.g., Gaussian beam). A comblike grating surface is considered in the analysis, for which a theoretical model for predicting the scattered fields has been developed. Numerical results on scattered surface wave as well as bulk waves are given for different parameters of the incident beam. It is found that at low frequencies ($ka \ll 1$, where a is the grating period and k is the wave number in air), it is the inhomogeneity of the source that results in surface wave propagation in the scattered waves. The analysis is also applicable to impedance grating structures.

11:00

4aPAb3. *In situ* measurements of the fragipan acoustic to seismic coupling signature. Wheeler Howard and Craig J. Hickey (Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677)

The phenomena of acoustic to seismic (A/S) coupling, observed and studied since the 1950s, has most recently been used to detect shallow buried objects [Sabatier and Xiang, J. Acoust. Soc. Am. **105**, 1383 (1999); **106**, 2143 (1999)] and monitor detonation of nuclear weapons [Orcutt, J. Acoust. Soc. Am. **105**, 1038 (1999)]. At an air–surface interface airborne acoustic energy is coupled into the ground as seismic energy. The ratio of the seismic and airborne waves constitutes the A/S coupling signature, which is distinctive to the underlying structure. Seismic energy received by a geophone at the interface contains information, via reflected waves, about the underlying subsurface layer, media, and boundaries. Of particular interest in the Mississippi River Valley is the fragipan layer. The fragipan is the layer that directly affects the growth of crops, rate of soil erosion, and rate of water absorption in underlying layers. In this presentation, the A/S coupling signature data taken at an agricultural field station and forward model are discussed.

11:15

4aPAb4. Acoustic-to-seismic coupling variations in cold regions. Donald G. Albert (USA CRREL, 72 Lyme Rd., Hanover, NH 03755, dalbert@crrel.usace.army.mil)

Experiments were conducted to investigate the variations that may occur in acoustic-to-seismic coupling arising from changes in local near-surface conditions. The emphasis of the investigations was on cold regions, where many different surface conditions exist and where conditions may change over a short time period from wind, precipitation, freezing, or thawing. The measurements were conducted by recording blank pistol shots with surface geophones and microphones. Results are presented for grassland, thin and thick seasonal snow covers, polar firn, thin grounded ice, thick glacier ice, and floating river ice. The ratio of induced ground motion to acoustic pressure ranged from 0.5 to 20 micro-meters per second per Pascal. Often two arrivals were detected on the geophones, a high-speed seismic compressional wave followed by the air wave. [Work funded by the U.S. Army.]

11:30

4aPAb5. Nonlinear acoustic techniques for landmine detection: Experiments and theory, Part II. Murray S. Korman (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402) and James M. Sabatier (Univ. of Mississippi, University, MS 38677)

A nonlinear acoustic technique for detecting buried landmines has been suggested by Donskoy [SPIE Proc. **3392**, 211 (1998); **3710**, 239 (1999)] and verified in a number of experiments by Sabatier *et al.* [J. Acoust. Soc. Am. **110**, 2757, 2758 (2001)]. Airborne sound involving two primary frequencies $f1$ and $f2$ undergo acoustic-to-seismic coupling which interacts with the compliant mine and soil to generate a difference frequency component that can effect the vibration velocity at the surface. An LDV profile across a buried mine at the difference frequency (chosen at a resonance) showed more relative sensitivity than a linear profile. A layer of soil in a large tube, terminated by an airbacked clamped thin elastic plate, is used to model nonlinear effects including Donskoy's ''bouncing'' soil–mine interface. The frequency spectrum of modes of the plate's response is investigated vs the amount of soil mass loading. In particular, the lowest mode frequency first decreases with loading and then increases. Linear response curves near resonance compare well with the difference frequency response as one primary is varied. [Work supported by the United States Army Communications-Electronics Command Night Vision and Electronic Sensors Directorate.]

11:45

4aPAb6. A fast effective fluid model for the scattering of normally incident sound off of a buried landmine. Doru Velea (Planning Systems, Inc., 12030 Sunrise Valley Dr., Reston Plaza I, Ste. 400, Reston, VA 20191), Roger Waxler (Univ. of Mississippi, University, MS 38677), and James M. Sabatier (Ft. Belvoir, VA 22060)

Landmines buried in the ground can be found acoustically by insonifying the ground and detecting a contrast between the vibratory motion of the ground surface directly above the mine and away from the mine. A model to predict the near-field scattering of low-frequency (~ 100 Hz) acoustic waves off of a landmine is presented. The mine is assumed to be a rigid cylinder with a compliant top. The ground (soil) is modeled as an effective fluid. Because of the short range of the scattered field at low frequencies it has been possible to replace a full space model with a waveguide model that is straightforward to implement numerically. The predictions of the model will be discussed. [Work supported by the U. S. Army Communications-Electronics Command Night Vision and Electronic Sensors Directorate and the U.S. Army Research Office.]

Session 4aPP

Psychological and Physiological Acoustics and Architectural Acoustics: Perceptual Effects in Real and Virtual Rooms

Barbara G. Shinn-Cunningham, Cochair
Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, Massachusetts 02215

Murray R. Hodgson, Cochair
Occupational Hygiene Program, University of British Columbia, 2206 East Mall, Vancouver, British Columbia V6T 1Z3, Canada

Chair's Introduction—9:15

Invited Papers

9:20

4aPP1. Perceptual effects in auralization of virtual rooms. Mendel Kleiner, Pontus Larsson, Daniel Vastfjall (Chalmers Rm. Acoust. Group, Chalmers Univ. of Tech., SE-41296 Gothenburg, Sweden), and Rendell R. Torres (Prog. in Architectural Acoust., Rensselaer Polytech. Inst., Troy, NY 12180-3590)

By using various types of binaural simulation (or ''auralization'') of physical environments, it is now possible to study basic perceptual issues relevant to room acoustics, as well to simulate the acoustic conditions found in concert halls and other auditoria. Binaural simulation of physical spaces in general is also important to virtual reality systems. This presentation will begin with an overview of the issues encountered in the auralization of room and other environments. We will then discuss the influence of various approximations in room modeling, in particular, edge- and surface scattering, on the perceived room response. Finally, we will discuss cross-modal effects, such as the influence of visual cues on the perception of auditory cues, and the influence of cross-modal effects on the judgement of ''perceived presence'' and the rating of room acoustic quality.

9:40

4aPP2. Effects of room reverberation on spatial release from masking. Richard L. Freyman (Dept. of Commun. Disord., Univ. of Massachusetts, Amherst, MA 01003) and Patrick M. Zurek (Sensimetrics Corp., Somerville, MA 02144)

When signal and masker sources are separated horizontally in an anechoic space, one of the primary benefits is that the acoustic shadow from the head leads to a substantial improvement in signal-to-noise ratio in the high frequencies at one of the ears, relative to conditions in which signal and noise are presented from a common location. This presentation will examine how the improved signal audibility created by head shadow is affected by room reverberation. Through room-simulation models, the effects of three variables will be examined: degree of horizontal separation, distance between the listener and the signal and masker, and degree of room reverberation. This paper will present an evaluation of how well the effects of these variables can be predicted by generalized room acoustics equations incorporating room volume and surface absorption. The reduction in high-frequency head-shadow effects in reverberant environments, combined with a disruption of cues for low-frequency binaural release from masking, severely reduces the benefits of spatially separating signal and noise. Fortunately, due to the precedence effect, the perceived locations of signal and noise are accurate even in a reverberant space. The intelligibility benefits of perceived differences in signal and masker locations will be discussed. [Work supported by NIDCD.]

10:00

4aPP3. Effects of reverberation on speech segregation. John F. Culling, Chaz Yee Toh, and Kathryn I. Hodder (School of Psych., Cardiff Univ., P.O. Box 901, Cardiff CF10 3YG, UK)

Perceptual separation of speech from interfering noise using binaural cues and fundamental frequency ($F0$) differences is disrupted by reverberation [Plomp, Acoustica **34**, 200–211; Culling *et al.*, Speech Commun. **14**, 71–96]. Culling *et al.* found that the effect of $F0$ differences on vowel identification was robust in reverberation unless combined with even subtle $F0$ modulation. In the current study, speech reception thresholds (SRTs) were measured against a single competing voice. Both voices were either monotonized or normally intonated. Each came from recordings of the same voice, but interfering sentences were feminized ($F0$ increased

80%; vocal-tract length reduced 20%). The voices were presented from either the same or from different locations within anechoic and reverberant virtual rooms of Culling *et al.* In anechoic conditions, SRTs were lower when the voices were spatially separated and/or intonated, indicating that intonated speech is more intelligible than monotonous speech. In reverberant conditions (T60=400 ms), SRTs were higher, with no differences between the conditions. A follow-up experiment introduced sentences with inverted $F0$ contours. While acceptable in quiet, these sentences gave higher SRTs in all conditions. It appears that reverberant conditions leave intonated speech intelligible, but make it unseparable, while monotonous speech remains separable but is unintelligible.

10:20–10:30 Break

10:30

4aPP4. Subjective assessment of listening environments in classrooms. Murray Hodgson and Susan Kennedy (SOEH, Univ. of British Columbia, 3rd Fl., 2206 East Mall, Vancouver, BC V6T 1Z3, Canada, hodgson@mech.ubc.ca)

A questionnaire was developed to evaluate subjective perception of the listening environment by university classroom users. The questionnaire was administered to over 5700 students in 30 randomly selected classrooms. The questionnaire recorded students' perceptions of the listening environment and possible modifying factors. A ''perception of listening ease'' (PLE) score was generated. Items most associated with a poor listening environment were noise from other students and intermittent noises outside the classroom. Over 50% of students reported frequent difficulty with hearing questions from other students. Poorer PLE score was associated with high background noise from students, an instructor who articulates poorly, poorer classroom lighting and ''room air,'' microphone use, and decreased room volume per student. In large classrooms, hearing-impaired students scored PLE lower. Results indicate that PLE is a reliable and valid measure of classroom users' subjective perceptions of the listening environment, and that optimal classroom acoustical design needs to take into consideration ''in-use'' conditions, as well as physical characteristics of the empty classroom.

Contributed Papers

10:50

4aPP5. Effects of natural and artificial spatialization cues on segregation. Alain de Cheveigné (CNRS, Ircam, 1 Pl. Igor Stravinsky, 75004 Paris, France, cheveign@ircam.fr), Reinhard Gretzky, Alexis Baskind, and Olivier Warusfel (Ircam, 1 Pl. Igor Stravinsky, 75004 Paris, France)

A series of experiments was performed to better understanding factors that determine the clarity or ''transparency'' of sound scenes, particularly those created by artificial means. Using a paradigm proposed by C. J. Darwin and R. W. Hukin [J. Exp. Psychol. **25**, 617–629], subjects were presented with a target sentence containing one word (chosen among two) that they had to report. Simultaneously they were presented with a distractor sentence containing a second word, temporally aligned with the first. In the absence of segregation cues, subjects scored 50% correct on average. A higher score indicated that simultaneous segregation and sequential grouping mechanisms were both effective. Stimuli were presented by headphones using individual or dummy-head head-related transfer functions (HRTFs), and various combinations of source positions, room effects, and restitution techniques. [Work supported by the Cognitique Programme of the French Ministry of Research and Technology.]

11:05

4aPP6. The influence of reverberation on spatial release of masking in consonant identification. Sasha Devore, Barbara G. Shinn-Cunningham, Nathaniel I. Durlach, and H. Steven Colburn (Hearing Res. Ctr., Boston Univ., 44 Cummington St., Boston, MA 02215, sashad@mit.edu)

This study investigates how reverberation influences spatial release of masking in consonant identification. HRTFs were measured in two rooms with different reverberation characteristics from sources directly in front of and to the right of the listener at a distance of 1 m. These two sets of room HRTFs and pseudo-anechoic HRTFs (time-windowing out the reverberation) were used to simulate a speech target and speech-shaped masker over headphones in two spatial configurations. In both configurations, the target was directly ahead of the listener; the masker was either directly ahead or 45° to the right of the listener. Subjects were asked to identify the initial or final consonant (one of 12 stop and fricative consonants) in CVC nonsense syllables for different target-to-masker levels. Each spatial configuration and room condition was tested using binaural, monaural-left, and monaural-right headphone presentations. Consonant confusion matrices are analyzed to determine how specific acoustic features of consonants in different frequency regions are affected by reverberation and noise. Comparisons across the different conditions reveal how binaural and spatial cues influence consonant identification and how different acoustic environments affect both monaural and binaural performance. [Work supported in part by AFOSR Grant Nos. F49620-01-1-0005 and F49620-98-1-0442.]

11:20

4aPP7. Effect of reverberation on speech intelligibility in young children. Ruth Litovsky (Waisman Ctr., Univ. of Wisconsin-Madison, 1500 Highland Ave., Madison, WI 53705, Litovsky@waisman.wisc.edu)

Children spend a majority of their time in noisy and reverberant environments, where a listener's abilities to understand speech and localize its source are compromised. The present study utilizes a novel test, which was developed in our lab for the purpose of assessing the ability of young children to understand speech in the presence of competing sources, and the extent to which they benefit from spatial separation between target speech and competitors. A 4AFC identification procedure was used, with (familiar) words from the Spondee List for targets. Testing was conducted in three rooms with different spatial dimensions and acoustic characteristics (reverberation times T60 equal to 250 ms, 500 ms, and 1200 ms). The target was always presented from the front. Conditions were: quiet, with competing sounds near the target, or with competing sounds 90° away from the target. Data were collected from children with normal hearing, ages 4–7 years. Results suggest that, while children benefit from spatial separation of target and competitors, this benefit becomes increasingly smaller as the amount of reverberation in the room increases. This finding may be relevant in our efforts to improve room acoustics and enhance children's ability to learn in classrooms.

Session 4aSA

Structural Acoustics and Vibration: Flow-Induced Vibration and Noise

Yan-Fan Hwang, Chair
Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

Chair's Introduction—8:00

Invited Papers

8:05

4aSA1. Prediction of flow induced sound: Past, present, and future. William Blake and Irek Zawadzki (David Taylor Model Basin, MacArthur Blvd., Bethesda, MD 20817)

The prediction of flow-induced vibration and sound has developed into a refined analysis technology. At its genesis the capability to provide engineering evaluations of flow-driven surfaces was crude and semi-empirical. The ability to provide acoustic estimations of fidelity depended on empiricism, similitude, and engineering experience. Very little was known of the physics of flow–structure interaction to permit otherwise. Currently, capabilities to make high-fidelity engineering predictions have benefitted from continual advances in three major areas. Our knowledge of the physics of flow sources has matured to understand acoustically-relevant flow structures and the interaction of flows with surfaces. Our ability to computationally model structural response and acoustic Green's functions has evolved because of parallel advances in structural acoustics. Computational fluid dynamics has developed into a refined tool for simulating flow over complex geometries. It is continuing to evolve as we learn how to model acoustically relevant subsonic flow structures with large eddy simulation and direct numerical simulation. Once principally a technology available only to military application, these tools are becoming more commonplace in industrial applications. This lecture will trace some of these developments, show some examples, and indicate promising areas of advancement.

8:30

4aSA2. Trailing edge noise reduction in a backward-curved impeller. Gerald C. Lauchle (Grad. Prog. in Acoust., Penn State Univ., 218-B Applied Science Bldg., University Park, PA 16802)

Motorized impellers are used in many air-moving applications including room circulation, duct flow, roof and wall exhaust, and cooling of electronic components in cabinets. These fans are backward-curved centrifugal blowers that operate with no volute casing. These fans radiate broadband noise due to turbulence ingestion and trailing edge (TE) noise generating mechanisms. Considered here are trailing edge noise generation and its reduction in a typical motorized impeller. The sound power of the subject fans is measured in an acoustically transparent test plenum according to ANSI Standard S12.11-1987. Two different serrated TE treatments are designed. The designs assume that a turbulent boundary layer exists at the blade TE, but the actual fan Reynolds number based on chord length is transitional. Therefore, to assure that a turbulent boundary layer exists at the TE, two different inlet turbulators are implemented. These trip the blade boundary layer to a turbulent state. Reported are the effects of the TE serrations and turbulators acting individually on the fan noise, along with the synergistic effects of using them in combinations. Up to 6 dBA of noise reduction is observed when the two are used together. [Work supported by Nortel Networks.]

8:55

4aSA3. Predicting wall pressure fluctuations beneath a turbulent flow. Theodore M. Farabee (Naval Surface Warfare Ctr., Carderock Div. Signatures Directorate, 9500 MacArthur Blvd., West Bethesda, MD 20817-5700)

The source of excitation for a wide range of flow-induced vibration and noise problems is the unsteady surface pressure field imposed by flow of a turbulent boundary layer (TBL). In addition, for flow-over discontinuities, the TBL surface pressures constitute an incident wave field that is scattered at discontinuities resulting in sound production for even a rigid surface. For these reasons alone, interest in predicting TBL wall-pressure fluctuations has existed as long as there has been an interest in understanding and estimating flow-induced noise and vibration. Current approaches for predicting, or estimating, wall-pressure fluctuations fall into one of three general categories: direct numerical computations, in which the governing equations are solved directly; a statistical modeling approach which couples time-averaged flow calculations to theoretically derived modeling functions for wall pressures; and empirical methods which assume that wall pressures follow a known scaling relationship formed using relatively standard flow parameters that are easy to prescribe for the flow of interest. The merits of using each of these methods for the purposes of assessing flow-induced noise and vibration will be reviewed and discussed.

9:20

4aSA4. Estimation of broadband power radiated from ribbed plates under turbulent boundary layer excitation. M. L. Rumerman (Signatures Directorate, Naval Surface Warfare Ctr. Carderock Div., 9500 MacArthur Blvd., West Bethesda, MD 20817-5700)

''Exact'' calculation of the acoustic power radiated due to the vibration of ribbed underwater panels driven by a turbulent boundary layer (TBL) requires consideration of the interactions among ribs and end supports. This paper shows that the power radiated in broad frequency bands may be directly estimated by considering the discontinuities to act independently. Analysis of a string loaded with identical masses or springs, and subjected to a wave-number-white forcing spectral density, is first used to demonstrate that the mean-squared attachment force within a broad frequency band may be estimated by ignoring coupling among attachments. A water-loaded membrane having parallel line mass attachments is then used to show that each attachment can be considered to radiate independently of the others when the acoustic wavelength is less than the spacing between them. These simple models are generalized and combined to develop an estimate of the vibration-related power radiated by finite ribbed steel plate sections in water driven by TBL-like surface pressures. The approximation is compared to results based on an exact formulation for a doubly periodic plate that represents ribs and end supports, and found to agree within 3 dB. [Work supported by ONR.]

9:45

4aSA5. Unit-cell finite element analysis of the response of complex structures to flow excitation. Nathan C. Martin II (BBN Technologies, 10 Fawcett St., Cambridge, MA 02138, nmartin@bbn.com) and Robert N. Dees (BBN Technologies, Old Mystic Mill, 11 Main St., Mystic, CT 06355)

Analytic estimation of the vibration response of rib-reinforced structures excited by turbulent boundary layer (TBL) wall pressure fluctuations is a subject of ongoing interest for numerous applications in airborne and underwater acoustics. Previous investigators have developed a variety of closed form and approximate analytic methods for estimating the response of structures of interest. Fluid-loading and rib-reinforcement effects make problems involving the response of sea-going vessels particularly challenging. Recent advances in finite element based methods and the enhanced capabilities of modern computer hardware now permit the application of the FEA method to such flow-induced vibration problems. This paper describes the application of the unit-cell capability in BBN's SARA-2D finite element code to evaluate the response of fluid-loaded, rib-reinforced cylinders to TBL excitation. An overview of the unit-cell modeling method is presented, and example results are provided to illustrate the effects of variations in rib and shell dimensions.

10:10–10:25 Break

10:25

4aSA6. On sound generation mechanism by a centrifugal blower. Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

Centrifugal blower noise has often been modeled as dipoles and quadrupoles to account for the effects of fluid–structure interaction and turbulence as the impeller rotates. However, many experimental results have shown that sound powers from centrifugal blowers increase with speed to the power of 4–6, which implies the existence of a monopole [Lighthill, Proc. R. Soc., Ser. A **222**, 564–587 (1952)]. This paper demonstrates that such a monopole indeed exists for a blower running inside a heating ventilation and air-conditioning (HVAC) unit of a vehicle. Tests indicate that this monopole is producible by a volumetric fluctuation due to an unsteady rotor. When the blower is operated at the voltage power input specified by the car manufacturer, the sound power increases with the speed to the power of 4. When the blower is installed on a stable shaft and running at the same voltage power input, the volumetric fluctuations are significantly reduced and the sound power increases with speed to the power of 6. This implies that the monopole sound has been effectively replaced by the dipole sound. Since dipole is less effective in generating sound at low speeds than monopole, eliminating rotor fluctuations can lower vehicle HVAC noise levels.

Contributed Papers

10:50

4aSA7. Excitation by flow over an obstructed opening. Paul J. Zoccola, Jr. (Carderock Div. Naval Surface Warfare Ctr., 9500 MacArthur Blvd., West Bethesda, MD 20817, ZoccolaPJ@nswccd.navy.mil)

The effect on flow-induced cavity resonance of the presence of an obstruction in the cavity opening is considered. The presence of a single obstruction or of a grid generally alters the flow so that the excitation occurs on the smaller length scale created by the obstructions. However, discussion of resonant excitation on the length scale encompassing the obstructions has not been found in the literature. For this study, measurements of cavity pressure due to flow over a cavity with an obstruction of varying dimensions in the opening were made. Measurements of the flow field around the single obstruction were also made. The cavity pressure measurements show that flow over an obstructed opening does result in the occurrence of classical resonant excitation at the large length scale. The frequency of the excitation and the amplitude of the response at the large length scale are reduced, depending on the dimensions of the obstruction. Flow field results show the effects that an obstruction has on the flow, including effects on the vortex convection velocity and the energy production distribution.

11:05

4aSA8. Vibrations of infinite plates and the mean value vibrations of finite plates excited by turbulent boundary layer flows. Stephen Hambric and Yun-Fan Hwang (ARL/Penn State Univ., P.O. Box 30, State College, PA 16804, sah@wt.arl.psu.edu)

The vibration response of a turbulent boundary layer (TBL) excited flat plate is analyzed using a finite-element model and infinite plate theory. For the finite-element models, discretization sufficient to resolve the convective fluctuations in the flow excitation field is used for the study. Clamped boundary conditions are assumed for the finite-element models and analyses are conducted at a variety of flow speeds and structural loss factors. Two equivalent TBL wall pressure excitation models are applied to the plates: (1) a modified Corcos cross-spectrum model for the finite-element models, and (2) a wave-vector-frequency spectrum model (the

Fourier transform of the modified Corcos cross spectra) for the infinite plate theory. The TBL wall pressure autospectrum is approximated using the model derived by Smolyakov and Tkachenko. The infinite plate predicts the mean value response of the finite plate very well for all speeds and loss factors considered, showing consistency with Skudrzyks mean value theory and its usefulness for generating sound and vibration estimates.

11:20

4aSA9. Turbulent boundary layer pressure fluctuations at large scales. Wendy Sanders, Carolyn Judge, Eric Winkel, Steven Ceccio (Univ. of Michigan, 1231 Beal Ave., Ann Arbor, MI 48109, wendusa@engin.umich.edu), David Dowling, and Marc Perlin (Univ. of Michigan, 1231 Beal Ave., Ann Arbor, MI 48109)

The pressure fluctuations underneath a turbulent boundary are an important excitation source for noise and vibration for both aircraft and ships. This presentation describes experimental results from a new study of flat-plate turbulent boundary layer pressure fluctuations in water at large scales and high Reynolds number. The experiments were performed at the U.S. Navy's William B. Morgan Large Cavitation Channel in Memphis, TN on a polished flat plate 3.05 m wide, 12.8 m long, and 0.18 m thick. Flow velocity, skin friction, surface pressure, and plate acceleration measurements were made at multiple downstream locations at flow speeds ranging from 0.5 m/s to 19 m/s for a Reynolds number (based on downstream distance) range of several million to 200 million. Dynamic surface pressures were recorded with 16 flush mounted pressure transducers forming an L-array with streamwise dimension of 0.264 m and cross-stream dimension of 0.391 m. Measured 99% boundary-layer thicknesses were typically of order 0.10 m. Results for spatial and temporal correlation functions, as well as auto- and cross-spectra are presented and compared with prior lower-Reynolds-number results using either inner or outer variable scaling. [Work sponsored by the Defense Advanced Research Projects Agency and ONR Code 333.]

11:35

4aSA10. Flow noise predictions of a submerged cylinder under turbulent boundary layer excitations. Kuangcheng Wu (Dept. of Signatures and Hydrodynamics, Northrop Grumman Newport News, 4101 Washington Ave., Newport News, VA 23607, wu_k@nns.com) and Nickolas Vlahopoulos (Univ. of Michigan, Ann Arbor, MI 48109)

The unsteady fluctuated pressure underneath turbulent boundary layers (TBL) is one of major noise sources in moving vehicles. Recently, discretized TBL forcing functions have been applied to planar structures in air [Y. F. Hwang and S. A. Hambric, Noise-Con, 2000; M. Allen and N. Vlahopoulos, Computers and Structures, 2000; M. Allen and N. Vlahopoulos, *Finite Elements in Analysis and Design*, 2001; M. Allen, R. Sbragio, and N. Vlahopoulos, AIAA J. 2001]. This paper discusses prediction of the flow-induced radiated noise and surface responses of a submerged hemisphere-capped cylindrical shell ($L/D=11$). The FEM/IFEM (infinite finite element method) approach is used to calculate structural acoustic transfer functions and to accurately account for the fluid loading effects. The effect on TBL due to the curvature of a cylinder is captured by utilizing the potential flow—boundary layer theory to determine key boundary layer parameters. Predictions of the surface intensity and far field responses are developed through stochastic analysis due to the natural of the TBL excitations. A MATLAB script is generated to determine the power spectral density of the responses. [Work supported by ONR Code 334.]

11:50

4aSA11. Hydrofoil near-wake sound sources at high Reynolds number. Dwayne A. Bourgoyne, Joshua M. Hamel, Carolyn Q. Judge (Univ. of Michigan, Mech. Eng. Rm. 2010, WE Lay Autolab, 1231 Beal Ave., Ann Arbor, MI 48109, dbourgoy@umich.edu), Steve L. Ceccio, and David R. Dowling (Univ. of Michigan, Ann Arbor, MI 48109)

An important hydroacoustic noise source from a fully submerged noncavitating hydrofoil is often the unsteady separated turbulent flow near its trailing edge. Here, hydroacoustic noise may be produced by boundary layer turbulence swept past and scattered from the foils trailing edge, and by coherent vortices formed in the foils near-wake. Such vortices may generate an energetic tonal component that rises above the broadband trailing-edge hydroacoustic noise. This presentation describes results of an experimental effort to identify and measure vortical flow features in the near-wake of a two-dimensional hydrofoil at chord-based Reynolds numbers ranging from 0.5 to 60 million. The experiments were conducted at the U.S. Navy's William B. Morgan Large Cavitation Channel with a test-section-spanning hydrofoil (2.1 m chord, 3.0 m span) at flow speeds from 0.25 to 18.3 m/s. Two trailing-edge shapes were investigated, and foil-internal accelerometers were used to monitor structural vibration. Velocity fluctuation spectra were measured in the foils near-wake with a two-component LDV system, and dynamic surface pressures were measured near the foils trailing edge with flush-mounted transducer arrays. Both indicate Reynolds number and trailing-edge shape-dependent vortex shedding. [Significant assistance provided by personnel from NSWC-CD. Work sponsored by Code 333 of ONR.]

12:05

4aSA12. Two structural intensity prediction methods in plates excited by turbulent boundary layers. Michael J. Daley and Stephen A. Hambric (ARL and The Grad. Prog. in Acoust., Penn State Univ., 217 Applied Science Blvd., University Park, PA 16802, mdaley@sabine.acs.psu.edu)

Structural intensity (S-I) fields may be used to identify energy flow paths through a vibrating structure, as well as energy source and sink regions. Boundary layer excitation of structures occurs in numerous aerospace and underwater applications. This study describes two methods of predicting S-I fields in structures excited by turbulent boundary layers. The first prediction method combines well known multiple-input/multiple-output system theory with finite differencing techniques to obtain S-I predictions. The second method uses an analytic computation technique. This technique combines response matrices between applied pressures and response stresses and velocities to form cross spectra matrices with units of power. These cross spectra matrices are proportional to S-I. The two prediction methods are described in detail and a set of case studies incorporating these methods is presented. The case studies are based on a simply supported plate with an attached point damper. The Corcos model of the boundary layer spatial wall pressure correlation is applied to the plate due to its simplicity. The case studies include intensity fields due to flows above, at, and below aerodynamic coincidence (the frequency at which a structural wavelength equals a convective wavelength). [Work supported by The Applied Research Laboratory, Penn State.]

12:20

4aSA13. Effect of mean flow on the analysis of noise and vibration in hydraulic lines. Mardi C. Hastings (Biomed. Eng. Ctr., Ohio State Univ., 1080 Carmack Rd., Columbus, OH 43210)

Recent reports in the literature conclude that convective terms due to mean flow should be included in the analysis to predict noise and vibration in steel hydraulic lines containing compliant hoses [M. L. Munjal and P. T. Thawani, Noise Control Eng. J. **45**, 235–242 (1997)]. In this study analytical predictions for noise and vibration were compared with experimental data obtained from a transmission line excited by a 10-vane pump source. The predictions were based on a transfer matrix analysis that included the fluid–structure interaction but neglected the convective terms associated with mean flow [M. C. Hastings and C.-C. Chen, J. Passenger Cars, SAE Trans. **102**, 1762–1767 (1994)]. Experimental data and analytical predictions were in excellent agreement. These results indicate that mean flow convective terms are negligible and do not need to be included in the analysis of noise and vibration occurring in relatively low-power hydraulic assist systems (such as automotive power steering). A comparison of analytical results with and without mean flow convective terms for higher-power hydraulic systems indicates when the convective terms are needed for an accurate prediction of noise and vibration.

Session 4aSC

Speech Communication: Speech Perception of Normal and Impaired Hearing (Poster Session)

Peggy B. Nelson, Chair

Communication Disorders, University of Minnesota, 115 Shevlin Hall, 164 Pillsbury Drive, Southeast, Minneapolis, Minnesota 55455

Contributed Papers

All posters will be on display from 8:30 a.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 12:00 noon. To allow for extended viewing time, posters will remain on display until 5:00 p.m.

4aSC1. Speech recognition in continuous and fluctuating maskers by listeners with normal and impaired hearing: Effects of presentation level. Van Summers and Michelle Molis (Army Audiol. & Speech Ctr., Walter Reed Army Medical Ctr., 6900 Georgia Ave. NW, Washington, DC 20307-5001)

Listeners with normal hearing show better speech recognition in the presence of fluctuating background sounds, such as a single competing voice, than in unmodulated noise at the same overall level. These performance differences are greatly reduced in listeners with hearing impairment who generally show little benefit from fluctuations in masker envelopes. If this lack of benefit is entirely due to elevated quiet thresholds and the resulting inaudibility of low-amplitude portions of signal+masker, hearing-impaired listeners should show an increasing benefit from masker fluctuations as the presentation levels increase. Normally hearing and hearing-impaired listeners were tested for sentence recognition at moderate and high presentation levels in a competing speech-shaped noise, in competing speech by a single talker, and in competing time-reversed speech by the same talker. Normal-hearing listeners showed more accurate recognition in the fluctuating maskers than in unmodulated noise with some evidence that modulated:unmodulated performance differences may decrease at high presentation levels. Hearing-impaired listeners showed more similar performance across maskers and presentation levels. The results suggest that audibility does not completely account for the group differences in performance with fluctuating maskers: supra-threshold processing differences between groups may also contribute to the disparity in performance. [Work supported by NIH.]

4aSC2. The time course of learning during a vowel discrimination task by hearing-impaired and masked normal-hearing listeners. Carrie Davis, Diane Kewley-Port, and Maureen Coughlin (Dept. of Speech & Hearing Sci., Indiana Univ., Bloomington, 200 S. Jordan Ave., Bloomington, IN 47405)

Vowel discrimination was compared between a group of young, well-trained listeners with mild-to-moderate sensorineural hearing impairment (YHI), and a matched group of normal hearing, noise-masked listeners (YNH). Unexpectedly, discrimination of $F1$ and $F2$ in the YHI listeners was equal to or better than that observed in YNH listeners in three conditions of similar audibility [Davis *et al.*, J. Acoust. Soc. Am. **109**, 2501 (2001)]. However, in the same time interval, the YHI subjects completed an average of 55% more blocks of testing than the YNH group. New analyses were undertaken to examine the time course of learning during the vowel discrimination task, to determine whether performance was affected by number of trials. Learning curves for a set of vowels in the $F1$ and $F2$ regions showed no significant differences between the YHI and YNH listeners. Thus while the YHI subjects completed more trials overall, they achieved a level of discrimination similar to that of their normal-hearing peers within the same number of blocks. Implications of discrimination performance in relation to hearing status and listening strategies will be discussed. [Work supported by NIHDCD-02229.]

4aSC3. Optimal sweep cycle for time-varying comb filters for binaural dichotic presentation to improve speech perception in sensorineural hearing impairment. Alice N. Cheeran, Prem C. Pandey, and Dakshayani S. Jangamashetti (EE Dept., IIT Bombay, Powai Mumbai-400 076, India, pcpandey@ee.iitb.ac.in)

In a previous investigation [P. C. Pandey *et al.*, J. Acoust. Soc. Am. **110**, 2705 (2001)], a scheme using binaural dichotic presentation was devised for simultaneously reducing the effect of increased temporal and spectral masking in bilateral sensorineural hearing impairment. Speech was processed by a pair of time-varying comb filters with passbands corresponding to cyclically swept auditory critical bands, with the objective that spectral components in neighboring critical bands do not mask each other and sweeping of filter passbands provides relaxation time to the sensory cells on the basilar membrane. Presently investigation is carried out to find the optimal value of the sweep cycle. Comb filters used were 256-coefficient linear phase filters, with transition crossovers adjusted for low perceived spectral distortion, 1 dB passband ripple, 30 dB stopband attenuation, and 78–117 Hz transition width. Acoustic stimuli consisted of swept sine wave and running speech from a male and a female speaker. Bilateral loss was simulated by adding broadband noise with constant short-time SNR. Listening tests with stimuli processed using sweep cycles of 10, 20, 40, 50, 60, 80, 100 ms indicated highest perceptual quality ranking for sweep cycle in the 40–60 ms range, with a peak at 50 ms.

4aSC4. Preliminary observations of infants' detection of backward masked tones. Lynne A. Werner and Heather K. Parrish (Dept. of Speech and Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105-6246, lawerner@u.washington.edu)

Backward masked thresholds appear to have a prolonged developmental course compared to other measures of auditory capacity, but backward masking has not been studied in infant listeners. The present study examined 7–9-month-old infants' detection of 20-ms, 1-kHz pure tones masked by a 50-ms, 2.5-kHz-lowpass noise. The interval between the offset of the tone and the onset of the masker was 0 ms. Both stimuli had 5-ms rise and fall times. The spectrum level of the masker was 30-dB SPL. Thresholds for the backward masked tone were determined adaptively for 6 young adults using a rule that converges on the 71% correct point on the psychometric function. The average threshold was 50-dB SPL (SD=12 dB). Thirteen infants were trained to respond to a backward masked 95-dB SPL tone, but not to the masker alone. These infants then completed 30 single-

interval test trials, with 15 no-signal trials and 15 signal trials with the tone fixed at 85-dB SPL and the same backward masker. Infants achieved a p(C)max of about 0.7 (SD=0.05) in this task. These results suggest that at this age infants' backward masked threshold is about 35 dB higher than the adults'.

4aSC5. Perceptual cue weighting of voiceless stop consonants as represented by mismatch negativity and P300. Cliff Franklin and Ashley Harkrider (Dept. of Audiol. and Speech Pathol., Univ. of Tennessee, 457 S. Stadium Hall, Knoxville, TN 37996)

Changes in high-frequency content form the basis of differentiating between the perception of voiceless stop consonants, /p/ and /t/. Behavioral studies show that two cues, the formant transition onset frequency and the relative amplitude of the burst component of the stop consonant, contribute to the perception of these consonants, and that the weight of these cues differs for normal-hearing versus hearing-impaired individuals. Our goal in the current study was to obtain a measure of cue weighting electrophysiologically in twelve normal-hearing adults and compare the results with those previously reported in behavioral studies. Two physiological indices of perceptual discrimination of acoustical changes, the attention-independent mismatch negativity (MMN) and the attention-dependent P300, were measured by varying each cue independently and in combination. This resulted in three MMN and three P300 responses per subject in response to (a) changing the formant transition cue with a constant, neutral relative amplitude, (b) changing the relative amplitude cue with a constant, neutral formant transition cue, and (c) changing both cues. Measures from individual and group mean waveforms will be used to determine cue weighting based on sum of square values derived from multivariate analysis of variance. Results will be discussed relative to behavioral perception and attention.

4aSC6. Telephone speech enhancement for hearing-impaired listeners using multi-channel compression. Harikrishna P. Natarajan, Ashok K. Krishnamurthy (Dept. of Elec. Eng., The Ohio State Univ., 2015 Neil Ave., Columbus, OH 43210, krishnamurthy.1@osu.edu), and Lawrence L. Feth (The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210)

Elderly listeners with sensorineural hearing impairment usually have difficulty with telephone communication if they do not use a hearing aid or telephone amplifier. In many cases, these devices may not be readily available or maybe uncomfortable to use. We describe here an alternative strategy to enhance the speech by pre-processing the signal before sending it over the telephone line. The pre-processing is based on a multi-channel compression algorithm that modifies the critical band spectrum of the speech signal to bring it within the dynamic hearing range of the listener, while also compensating for the effects of the telephone line. We describe the algorithm and the results of a simulation study in which the Articulation Index (AI) is used to measure the intelligibility enhancement, and the Glasberg, Moore and Stone (1999) model is used to simulate the effects of the hearing loss.

4aSC7. On learning to recognize spectrally reduced speech: I. The effects of multiple repetitions. Michelle Eng (Dept. of Psych., Univ. of Cincinnati, ML0376, Cincinnati, OH 45221-0376) and Peter Chiu (Univ. of Cincinnati, Cincinnati, OH 45221-0376)

In a previous study [Chiu *et al.*, J. Acoust. Soc. Am. **109**, 2501 (2001)], normal listerers learned open-set recognition of words in Harvard IEEE sentences that were spectrally reduced to 4 channels using the CIS processing strategy [Shannon *et al.*, J. Acoust. Soc. Am. **104**, 2467–2476 (1998)]. Half of the listeners listened to a male talker during training, the other half a female talker. At different points in the training listeners were tested with blocks of IEEE sentences spoken by the familiar talker, blocks of IEEE sentences by the novel talker, blocks of IEEE sentences of the 2 talkers randomly intermixed, and blocks of HINT sentences. In that study, listeners could listen to each sentence during early training for an unrestricted number of times before they were given feedback. In the current study, listeners were allowed to listen to each sentence only once during training before feedback was given. Preliminary data obtained from the current study were largely comparable to those obtained in the previous study. The results suggest that massed repetitions and drills, at least in the context of our experimental conditions, wa have minimal effect on learning to recognize spectrally reduced speech. [Work supported by URC, Univ. of Cincinnati.]

4aSC8. The effects of room acoustics and multiple talkers on speech intelligibility with normal hearing and hearing impaired individuals. Pamela Mishler (Dayton Veterans Affairs Medical Ctr., pamela.mishler@wright.edu), Mark Ericson (Wright–Patterson Air Force Base, Dayton, OH), and Shawn Cowell (Dayton Affairs Medical Ctr., Dayton, OH)

Many studies have shown the effect of hearing loss and spatial separation on multi-talker word identification ability in free-field listening conditions. However, little is known about the combination of room reflections and hearing impairment on understanding speech. The purpose of this study was to determine if the effect of room reflections on speech intelligibility was equal among normal hearing and hearing impaired listeners. Two acoustic generation techniques were employed to create the free-field and semi-reverberant listening conditions. Free-field listening conditions were created using virtual acoustic technology. Room reflections were added to the speech samples using dummy-head recordings in a semi-reverberant room. The word identification test used in the experiments was the Modified Rhyme Test (MRT), which afforded a multitude of phonemic pairs when phrases were played simultaneously. Thirty-three listeners in each group participated in each test condition. The hearing impaired listeners were below 60 years of age and exhibited a high frequency, cochlear hearing loss. The normal hearing listeners were age matched with the hearing impaired group. A difference in the type of phonemic identifications between the two techniques and the two groups were found. These phonemic contrasts between will be discussed individually.

4aSC9. The role of reverberation in release from masking due to spatial separation of sources in speech recognition. Gerald Kidd, Jr., Tanya L. Arbogast, Andrew Brughera, Christine R. Mason (Dept. of Commun. Disord. and Hearing Res. Ctr., Boston Univ., 635 Commonwealth Ave., Boston, MA 02215, gkidd@bu.edu), and William M. Hartmann (Boston Univ., Boston, MA)

Arbogast *et al.* [ARO Mtg. (2002)] found a large release from masking obtained by spatial separation of a target talker and competing speech masker. Both stimuli were sentences from the Coordinate Response Measure corpus [Bolia *et al.*, J. Acoust. Soc. Am. (2000)] processed by extracting the envelopes of 15 narrow frequency bands and using the envelopes to modulate carrier tones at the center of each band. By playing nonoverlapping subsets (6–8) of bands from signal and masker they minimized the energetic component while maximizing the informational component of masking. This study extends that work to determine the interaction between reverberation, masker type, and spatial release from masking. Stimuli were processed and presented as above. The target sentence was played at 0-deg azimuth while the masker sentence was played at 0 or 90-deg azimuth. Noise–masker controls were also tested. The listening environment was an IAC booth having dimensions of 12 ft×13 ft. Acoustic extremes were achieved using Plexiglas™ (highly reflective) or foam (highly absorptive). The results indicated that the amount of masking and the spatial release from masking depend both on the characteristics of the room and masker type. Discussion will center on the acoustic and perceptual factors affecting performance. [Work supported by NIH/NIDCD.]

4aSC10. On learning to recognize spectrally reduced speech: II. Individual differences. Peter Chiu (Dept. of Psych. and Commun. Sci. and Disord., ML0376, Univ. of Cincinnati, Cincinnati, OH 45221-0376), Michelle Eng, Bethany Strange (Univ. of Cincinnati, Cincinnati, OH 45221-0376), Sasha Yampolsky, and Gloria Waters (Boston Univ., 635 Commonwealth Ave., Boston, MA 02215)

When patients with cochlear implants attempt speech recognition, or when listeners with normal hearing try to recognize spectrally reduced speech, performance may suffer because of additional demands on cognitive resources imposed by sensory degradation. Some previous studies have reported a significant correlation between digit span and recognition of spectrally reduced speech, but others have not [Eisenberg *et al.,* J. Acoust. Soc. Am. **107**, 2704–2710 (2000)]. In this study, we administered 3 separate measures of working memory capacity (i.e., digit span, alphabet span, and sentence span) and 1 measure of reading ability (American National Adult Reading Test, ANART) to normal-hearing listeners who participated in our previous studies [Chiu *et al.,* J. Acoust. Soc. Am. **109**, 2501 (2001) and Fishbeck and Chiu, J. Acoust. Soc. Am. **109**, 2504 (2001)]. Preliminary data show that the listener's ability to recognize spectrally reduced speech shows the strongest relationship with ANART scores and the weakest relationship with digit span. Implications of the current findings on theories of speech recognition and working memory will be discussed. [Work supported by URC, Univ. of Cincinnati.]

4aSC11. Role of auditory feedback in speech produced by cochlear implanted adults and children. Sneha V. Bharadwaj, Emily A. Tobey, Peter F. Assmann, and William F. Katz (Callier Ctr. for Commun. Disord., Univ. of Texas at Dallas, 1966 Inwood Rd., Dallas, TX 75235, snehab@utdallas.edu)

A prominent theory of speech production proposes that speech segments are largely controlled by reference to an internal model, with minimal reliance on auditory feedback. This theory also maintains that suprasegmental aspects of speech are directly regulated by auditory feedback. Accordingly, if a talker is briefly deprived of auditory feedback speech segments should not be affected, but suprasegmental properties should show significant change. To test this prediction, comparisons were made between speech samples obtained from cochlear implant users who repeated words under two conditions (1) implant device turned ON, and (2) implant switched OFF immediately before the repetition of each word. To determine whether producing unfamiliar speech requires greater reliance on auditory feedback than producing familiar speech, English and French words were elicited from English-speaking subjects. Subjects were congenitally deaf children ($n=4$) and adventitiously deafened adults ($n=4$). Vowel fundamental frequency and formant frequencies, vowel and syllable durations, and fricative spectral moments were analyzed. Preliminary data only partially confirm the predictions, in that both segmental and suprasegmental aspects of speech were significantly modified in the absence of auditory feedback. Modifications were greater for French compared to English words, suggesting greater reliance on auditory feedback for unfamiliar words. [Work supported by NIDCD.]

4aSC12. Spectral integration of synchronous and asynchronous cues to consonant identification. Joseph W. Hall III, Emily Buss, and John H. Grose (Univ. of North Carolina at Chapel Hill, Dept. of Otolaryngol., CB7070, Chapel Hill, NC 27599-7070)

When speech is presented at poor speech-to-masker ratios, the listener may momentarily receive some spectral regions of the signal at more favorable ratios than other regions. These optimal spectral regions may vary dynamically as a function of time. It is therefore of interest to determine how well listeners utilize cues from different spectral regions that occur asynchronously as opposed to synchronously across frequency. Here, we examined masked identification of consonants under various conditions of 10-Hz square-wave modulation of the speech material. Speech from 100–10 000 Hz was divided into either 2, 4, 8, or 16 contiguous bands. In one set of conditions, bands were amplitude modulated coherently, while in the other, even and odd bands were modulated 180 degrees out of phase. These conditions, along with further conditions employing only modulated even or odd bands, allowed performance to be compared between synchronous and asynchronous presentation. This approach is similar to a published checkerboard masking study in which the noise rather than the speech was modulated [Howard-Jones and Rosen, J. Acoust. Soc. Am. **93**, 2915–2922]. Preliminary data suggest that many listeners are able to utilize asynchronous speech cues efficiently in all of the band number conditions examined. [Work supported by NIDCD (5R01DC000418).]

4aSC13. Relationships among vocabulary size, nonverbal cognition, and spoken word recognition in adults with cochlear implants. Elizabeth A. Collison, Benjamin Munson, and Arlene E. Carney (Dept. of Commun. Disord., Univ. of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr., SE, Minneapolis, MN 55455, coll0301@umn.edu)

Recent research has attempted to identify the factors that predict speech perception performance among users of cochlear implants (CIs). Studies have found that approximately 20%–60% of the variance in speech perception scores can be accounted for by factors including duration of deafness, etiology, type of device, and length of implant use, leaving approximately 50% of the variance unaccounted for. The current study examines the extent to which vocabulary size and nonverbal cognitive ability predict CI listeners' spoken word recognition. Fifteen postlingually deafened adults with nucleus or clarion CIs were given standardized assessments of nonverbal cognitive ability and expressive vocabulary size: the Expressive Vocabulary Test, the Test of Nonverbal Intelligence-III, and the Woodcock-Johnson-III Test of Cognitive Ability, Verbal Comprehension subtest. Two spoken word recognition tasks were administered. In the first, listeners identified isophonemic CVC words. In the second, listeners identified gated words varying in lexical frequency and neighborhood density. Analyses will examine the influence of lexical frequency and neighborhood density on the uniqueness point in the gating task, as well as relationships among nonverbal cognitive ability, vocabulary size, and the two spoken word recognition measures. [Work supported by NIH Grant P01 DC00110 and by the Lions 3M Hearing Foundation.]

4aSC14. Speech feature discrimination in deaf children following cochlear implantation. Tonya R. Bergeson, David B. Pisoni, and Karen Iler Kirk (Indiana Univ. School of Medicine, Dept. of Otolaryngol., 699 West Dr. RR044, Indianapolis, IN 46202, tbergeso@iupui.edu)

Speech feature discrimination is a fundamental perceptual skill that is often assumed to underlie word recognition and sentence comprehension performance. To investigate the development of speech feature discrimination in deaf children with cochlear implants, we conducted a retrospective analysis of results from the Minimal Pairs Test (Robbins *et al.,* 1988) selected from patients enrolled in a longitudinal study of speech perception and language development. The MP test uses a 2AFC procedure in which children hear a word and select one of two pictures (bat–pat). All 43 children were prelingually deafened, received a cochlear implant before 6 years of age or between ages 6 and 9, and used either oral or total communication. Children were tested once every 6 months to 1 year for 7 years; not all children were tested at each interval. By 2 years postimplant, the majority of these children achieved near-ceiling levels of discrimination performance for vowel height, vowel place, and consonant manner. Most of the children also achieved plateaus but did not reach ceiling performance for consonant place and voicing. The relationship between speech feature discrimination, spoken word recognition, and sentence comprehension will be discussed. [Work supported by NIH/NIDCD Research Grant No. R01DC00064 and NIH/NIDCD Training Grant No. T32DC00012.]

4aSC15. Understanding speech in single-talker interference: Normal-hearing listeners and cochlear implant users. Peggy B. Nelson and Su-Hyun Jin (Dept. of Commun. Disord., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455)

Our previous data [Nelson *et al.*, ASA (2001)] indicated that users of cochlear implants experience little release from masking when listening to speech in modulated maskers. Although normal-hearing listeners take advantage of temporal dips in modulated maskers, cochlear implant users do not. The current study investigated listeners with normal hearing and cochlear implants for their understanding of speech in the presence of a single-talker masker. Participants listened to IEEE sentence material spoken clearly at a slow rate (3.5 syllables per second) by either a male (average $F0=110$ Hz) or a female (average $F0=210$ Hz) talker. Maskers included continuous speech from other male and female talkers spoken at both slow and fast rates. Results from normal-hearing listeners indicated that when the signal and masker had a similar $F0$ and rate, identification was poorer than when signal and masker differed in $F0$ and/or rate. The primary factor affecting performance was the $F0$ of the masker. It is hypothesized that cochlear implant users will not show the effect of talker/masker $F0$, but will show decreased performance when talker and masker have the same rate. Results from cochlear implant listeners will be reported. [Work supported by NIDCD P01DCD00110.]

4aSC16. Perception of talker similarity by normal-hearing children and hearing-impaired children with cochlear implants. Miranda Cleary (Speech Res. Lab., Dept. of Psych., Indiana Univ., 1101 E. Tenth St., Bloomington, IN 47405)

This project investigated judgments of perceptual similarity, specifically, how similar do two sentences need to be in terms of fundamental and formant frequencies for children to categorize both utterances as spoken by the same talker? Same–different judgments were obtained using an adaptive testing procedure in which the similarity of voice pairs was systematically varied. Participants included five-year-old normal-hearing children and prelingually deafened school-age cochlear implant users with 2+ years of implant experience. Stimuli consisted of natural sentence-length utterances of a fixed syntactic form processed and resynthesized to form a stimulus continuum of differentially similar-sounding voices. Auditory comprehension skills were measured using materials from the same stimulus set. Comprehension was also measured against a competing talker masker. An adaptive testing procedure adjusted the similarity of masker voice to target voice according to each child's performance. The probability of ''same talker'' judgments is reported for each group of children as a function of pitch difference size. Preliminary results indicate a commonality in the auditory skills necessary for both talker discrimination and auditory comprehension. The relationship between discrimination of similar-sounding talkers and the ability to perceptually separate voices will also be discussed. [Work supported by NIH-NIDCD.]

4aSC17. Auditory streaming in cochlear implant listeners. Monita Chatterjee and John J. Galvin III (Dept. of Auditory Implants and Percept., House Ear Inst., 2100 W. Third St., Los Angeles, CA 90057)

We used the following method to measure auditory streaming in CI listeners: a ''test'' pattern is composed of two loudness-matched tones A and B presented in the sequence AABAABAAB . . . . If A and B are perceptually separable, the subject should hear ''AA . . . AA . . . AA . . . '' and ''B . . . B . . . B . . . '' as the two separate streams. The subject hears a ''preview'' sequence of tones which can have the rhythm of either tone B or tone A in the pattern (e.g., X . . . X . . . X . . . X or XX . . . XX . . . XX . . .), where X is the preview tone. The subject indicates whether or not she/he heard the rhythm of the preview sequence in the test sequence. A number of sets of A, B, and X are set up, and the subject is presented with the different patterns in random order. Results suggest CI listeners are able to perceptually separate tonal sequences into auditory streams. The tonotopic distance (electrode pair separation) between tones is an important factor for streaming. However, listeners are also able to stream sequences of tones presented to the same electrode pair (tonotopic location) but with different temporal envelopes. [Work supported by NIDCD.]

4aSC18. Cochlear-implant simulations of speech in complex acoustic backgrounds. Michael K. Qin and Andrew J. Oxenham (MIT, 77 Massachusetts Ave., Cambridge, MA 02139-4307)

Cochlear-implant simulations have shown that good speech intelligibility in quiet can be obtained with very poor spectral resolution and no temporal fine-structure information. It is not known how robust speech is to such processing in the presence of complex interfering sounds. This study examined speech reception for sentences in the presence of four types of interference (broadband noise, broadband noise modulated with a speech envelope, a single male talker, and a single female talker) and three processing schemes (24- and 8-channel cochlear-implant simulations and unprocessed). The spectral resolution in the 24-channel condition was designed to be comparable to that of normal hearing, but the original temporal fine structure was replaced by noise in both the 24- and 8-channel processing conditions. Performance with 24-channel processing was substantially poorer than in unprocessed conditions, despite the relatively good frequency resolution. Eight-channel processing produced the poorest performance. Stimulus processing was generally more detrimental to speech reception in the fluctuating interferers than in the unmodulated noise. The detrimental effects of eliminating fine-structure information may be due to the reduction in pitch cues, which listeners may use to help segregate the sources. [Work supported by NIDCD Grant R01DC05216.]

4aSC19. The contribution of auditory temporal processing to the separation of competing speech signals in listeners with normal hearing. Trudy J. Adam and Kathy Pichora-Fuller (Audiol. & Speech Sci., Univ. of BC #226, 5804 Fairview Ave., Vancouver, BC V6T 1Z3, Canada)

The hallmark of auditory function in aging adults is difficulty listening in a background of competing talkers, even when hearing sensitivity in quiet is good. Age-related physiological changes may contribute by introducing small timing errors (jitter) to the neural representation of sound, compromising the fidelity of the signal's fine temporal structure. This may preclude the association of spectral features to form an accurate percept of one complex stimulus, distinct from competing sounds. For simple voiced speech (vowels), the separation of two competing stimuli can be achieved on the basis of their respective harmonic (temporal) structures. Fundamental frequency ($F0$) differences in competing stimuli facilitate their segregation. This benefit was hypothesized to rely on the adequate temporal representation of the speech signal(s). Auditory aging was simulated via the desynchronization ($\sim$0.25-ms jitter) of the spectral bands of synthesized vowels. The perceptual benefit of $F0$ difference for the identification of concurrent vowel pairs was examined for intact and jittered vowels in young adults with normal hearing thresholds. Results suggest a role for reduced signal fidelity in the perceptual difficulties encountered in noisy everyday environments by aging listeners. [Work generously supported by the Michael Smith Foundation for Health Research.]

4aSC20. Identification of musical instruments by normal-hearing subjects listening through a cochlear-implant simulation. Rebecca D. Reich (MIT Media Lab., E15-492 20 Ames St., Cambridge, MA 02139, rreich@mit.edu) and Donald Eddington (Cochlear Implant Res. Lab., Boston, MA 02114)

Signal processing in a cochlear implant (CI) is primarily designed to convey speech and environmental sounds, and can cause distortion of musical timbre. Systematic investigation of musical instrument identification through a CI has not yet revealed how timbre is affected by the implant's processing. In this experiment, the bandpass filtering, rectification, and low-pass filtering of an implant are simulated in MATLAB. Syn-

thesized signals representing 12 common instruments, each performing a major scale, are processed by simulations using up to 8 analysis channels. The unprocessed recordings, together with the 8 simulation conditions for 12 instruments, are presented in random order to each of the subjects. The subject's task is to identify the instrument represented by each item. The subjects also subjectively score each item based on similarity and pleasantness. We anticipate performance using the simulation will be worse than the unprocessed condition because of the limited information delivered by the envelopes of the analysis channels. These results will be analyzed as a confusion matrix and provide a basis for contrasting the information used by subjects listening to the unprocessed and processed materials. Understanding these differences should aid in the development of new processing strategies to better represent music for cochlear implant users.

4aSC21. Typicality ratings of male and female voices. Brian Spisak, John Mullennix, Kelly Moro, Jessica Will (Dept. of Psych., Univ. of Pittsburgh, Johnstown, Johnstown, PA 15904), and Lynn Farnsworth (Wayne State Univ., Detroit, MI 48202)

Researchers have suggested that human voices are represented in memory in terms of prototypes [e.g., Kreiman and Papcun (1991); Papcun *et al.* (1989)]. Others have suggested that speech utterances are stored in memory via detailed exemplar-based representations [e.g., Lachs *et al.* (2000)]. The goal of the present study was to provide the first step toward assessing the viability of a prototype view of voice. Ten hVd utterances were recorded from each of 20 male and 20 female speakers. The utterances were blocked by speaker gender and presented to male and female listeners who rated each stimulus on a 1–7 typicality scale from ''least typical voice'' to ''most typical voice.'' There were significant effects of the type of vowel and speaker voice on the ratings, as well as interactions of vowel type with gender of subject and speaker voice. The results are discussed in terms of the strength of evidence for a graded category structure of voice categories that would be consistent with a prototype perspective of long-term memory representations of voice.

4aSC22. Study of effect of low-pass filtering for channels amplitude estimation on speech intelligibility by acoustic simulation of SPEAK strategy. Erdenebat Dashtseren and Shigeyoshi Kitazawa (Grad. School of Sci. and Eng., Johoku 3-5-1 Hamamatsu, Shizuoka 432-8011, Japan, bat@cs.inf.shizuoka.ac.jp)

The conventional way of amplitude envelope estimation for stimulation pulse train generation for selected channels within multi-channel cochlear implant speech processors is low-pass filtering of half or full wave rectified (HWR or FWR) frequency bands of the corresponding channels. To study the necessity of low-pass filtering (LPF) for amplitude envelope generation, acoustic simulation of the SPEAK speech strategy was implemented. Stimulation pulse patterns for selected channels are estimated without LPF and compared to those estimated by the conventional way. Acoustic stimuli are regenerated using an overlap adding technique of filtered pulse trains. Speech intelligibility tests on CV and VCV of Japanese within normal-hearing subjects were conducted to measure the effect of LPF. A one-way ANOVA indicates that the LPF with a cutoff frequency 160 Hz for amplitude envelope estimation of selected channels negatively affected the speech intelligibility.

4aSC23. Further investigations on binaural intelligibility of reverberant speech. Brad Libbey and Peter Rogers (Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405, gt1556a@prism.gatech.edu)

There is a binaural intelligibility advantage for many subjects when listening to reverberant speech. Two processes offer explanations as to why the binaural advantage exists. First, directional information may identify echoes and remove a portion of the reverberation in a manner similar to the precedence effect. A more likely possibility is that the reverberation behaves as additive noise with a specific time and frequency dependence. In this case, subjects may utilize the coherence of the left and right channels to limit the effects of the incoherent reverberation, similar to binaural masking release. To test the latter hypothesis, subjects listen to diotic (L+R), binaural (L&R), and diotic (L or R) reverberant words. The diotic treatment (L+R) adds two recordings measured at the location of the listener's ears. Since the listener faces the source, the direct signals are coherent while some cancellation is expected for the reverberation. The presentation will compare diotic (L+R) intelligibility to the binaural (L&R) condition as well as the diotic (L or R) control condition where the left or right signal plays to both ears.

4aSC24. The perception of voice pitch by naïve and experienced listeners. Donald Fucci (School of Hearing, Speech and Lang. Sci., Ohio Univ., Grover Ctr. W222, Athens, OH 45701, fucci@ohio.edu), Marianne Schueller, and Zinny Bond (Ohio Univ., Athens, OH 45701)

Perceptual judgments of a single speaker's voice pitch will be characterized by any given listener along a subjective continuum. Individual listener attributes have been suggested as an explanation for this diversity in categorizing voice pitch. Professionals involved in the evaluation of voice, specifically voice pitch, are expected to possess a level of perceptual skill that reliably quantifies pitch regardless of personal attributes. Two groups of individuals, one naïve and one experienced, were asked to perceptually match voice pitch during two pitch matching tasks. The first task involved matching an audio taped voice pitch produced on the vowel /a/ to a musical note on a keyboard. The second task involved listening to an audio-taped voice producing words, then matching the voice pitch of the word to a musical note on a keyboard. Accuracy of perceptual judgments was measured upon stimuli produced and assigned musical keyboard note by master's level opera music majors. Results indicate that a difference exists between groups in only one of the four variables measured. Individual differences varied, but appeared unrelated to previous musical experience.

4aSC25. Peripheral and central locus of a nonspeech phonetic context effect. Sarah C. Sullivan and Andrew J. Lotto (Washington State Univ., Pullman, WA 99164-4820, alotto@wsu.edu)

Previous work has demonstrated that nonspeech sounds with the appropriate spectral characteristics can affect the identification of speech sounds [Lotto and Kluender, Percept. Psychophys. **60**, 602–619 (1998)]. It has been proposed that these spectral context effects are due to interactions in the peripheral auditory system. For example, they could be the result of masking at the auditory nerve or of auditory enhancement effects that have been demonstrated to be monaural [Summerfield and Assmann, Percept. Psychophys. **45**, 529–536 (1989)]. To examine the locus of the context effect, synthesized syllables varying from /da/ to /ga/ were preceded by single-formant stimuli that mimicked the third formant of the syllables /al/ and /ar/. The nonspeech stimulus was presented either to the same or opposite ear as the target speech stimulus. Subjects speech identifications were shifted as a function of context in predicted directions for both presentation conditions. However, the size of the shift was smaller when the context was in the ear contralateral to the target syllable. These results agree well with similar results for speech contexts. The data suggest that the context effects occur at multiple levels of the auditory system and are not simply examples of masking or auditory enhancement.

4aSC26. Reverberation reduction by sub-band unmasking. Brad Libbey and Peter Rogers (Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405, gt1556a@prism.gatech.edu)

Reverberation reduces the intelligibility of speech recorded in rooms. Signal processing algorithms attempt to reduce this reverberation such that the speech not only sounds less reverberant, but is more intelligible. His-

torically, these techniques are based on calculations of the room impulse response, on arrays of two or more microphones, noise cancellation, or cepstral techniques. The new technique assumes that the intelligibility loss is due to overlap masking of reverberant tails of speech. The algorithm looks at sub-bands and identifies those that mask adjacent bands. When this occurs in conjunction with a decaying envelope the decay rate is increased. Thus, the masked band has a greater chance of being heard. Intelligibility results for the processed speech will be presented.

4aSC27. Effect of preceding speech on nonspeech sound perception. Joseph D. Stephens and Lori L. Holt (Psych. Dept., Carnegie Mellon Univ., and Ctr. for the Neural Basis of Cognition, 5000 Forbes Ave., Pittsburgh, PA 15213)

Data from Japanese quail suggest that the effect of preceding liquids (/l/ or /r/) on response to subsequent stops (/g/ or /d/) arises from general auditory processes sensitive to the spectral structure of sound [A. J. Lotto, K. R. Kluender, and L. L. Holt, J. Acoust. Soc. Am. **102**, 1134–1140 (1997)]. If spectral content is key, appropriate nonspeech sounds should influence perception of speech sounds and vice versa. The former effect has been demonstrated [A. J. Lotto and K. R. Kluender, Percept. Psychophys. **60**, 602–619 (1998)]. The current experiment investigated the influence of speech on the perception of nonspeech sounds. Nonspeech stimuli were 80-ms chirps modeled after the $F2$ and $F3$ transitions in /ga/ and /da/. $F3$ onset was increased in equal steps from 1800 Hz (/ga/ analog) to 2700 Hz (/da/ analog) to create a ten-member series. During AX discrimination trials, listeners heard chirps that were three steps apart on the series. Each chirp was preceded by a synthesized /al/ or /ar/. Results showed context effects predicted from differences in spectral content between the syllables and chirps. These results are consistent with the hypothesis that spectral contrast influences context effects in speech perception. [Work supported by ONR, NOHR, and CNBC.]

4aSC28. Recognition masking for transitions in initial and final position. Jeremy R. Gaston and Richard E. Pastore (Dept. of Psych., Binghamton Univ., P.O. Box 6000, Binghamton, NY 13901)

There is a common implicit assumption that the input to higher or specialized processes is accurately represented by the physical properties of the stimulus, with, at most, only detection masking and some basic initial filtering. Since playing an auditory stimulus backwards does not alter the spectral composition of the stimuli, this initial assumption has led to the conclusion that any differences in perception must be attributed to higher level or specialized processes. Thus, for example, significant differences in the perception of temporally reversed CV syllables are assumed to reflect only specialized processes (i.e., speech module). The current study challenges this basic implicit assumption. Backward ''recognition'' masking is a basic, but poorly understood perceptual phenomenon in which an added following stimulus interferes with discrimination of simple preceding tones. Recent work from our laboratory indicates similar interference in the recognition of easily discriminated chords. The current study investigates the role of backward recognition masking in determining ability to recognize the direction of frequency glides similar to formant transitions of stop consonants.

4aSC29. Acoustical considerations for an automotive hands-free mobile terminal. Richard J. Ruhala and Vasu Iyengar (Agere Systems, 1247 S. Cedar Crest Blvd., Allentown, PA 18103, ruhala@agere.com)

The harsh acoustical environment of an automotive interior yields significant difficulty in controlling acoustic echo and noise during ''hands-free'' wireless phone calls. In addition to having multiple closely coupled reflection paths, the user (talker) is often positioned between the loudspeaker and microphone path. This can cause frequent changes in the acoustic path which necessitates fast and accurate adaptation of the acoustic echo canceller and/or non-linear controller. Further, the frequency response between the talker and the microphone is altered due to cabin resonances and modes; it is also strongly affected by their positions. Additional acoustical considerations for automotive hands-free implementations are the high levels and non-stationary statistics of background noise, high amplification, and possible resultant distortion of the received audio signal, and non-linear network echo. These impairments yield significant challenges in controlling the acoustic echo and noise in full-duplex, partial-duplex, and even half-duplex systems. In this work, impulse response, frequency response, and SNR measurements are presented and analyzed to demonstrate the critical differences between a traditional desktop speakerphone and an automotive hands-free mobile terminal.

4aSC30. Acoustic and perceptual effects of overall $F0$ range in a lexical pitch accent distinction. Travis Wade (Dept. of Linguist., Univ. of Kansas, 1541 Lilac Ln., 427 Blake Hall, Lawrence, KS 66044-3177)

A speaker's overall fundamental frequency range is generally considered a variable, nonlinguistic element of intonation. This study examined the precision with which overall $F0$ is predictable based on previous intonational context and the extent to which it may be perceptually significant. Speakers of Tokyo Japanese produced pairs of sentences differing lexically only in the presence or absence of a single pitch accent as responses to visual and prerecorded speech cues presented in an interactive manner. $F0$ placement of high tones (previously observed to be relatively variable in pitch contours) was found to be consistent across speakers and uniformly dependent on the intonation of the different sentences used as cues. In a subsequent perception experiment, continuous manipulation of these same sentences between typical accented and typical non-accent-containing versions were presented to Japanese listeners for lexical identification. Results showed that listeners' perception was not significantly altered in compensation for artificial manipulation of preceding intonation. Implications are discussed within an autosegmental analysis of tone. The current results are consistent with the notion that pitch range (i.e., specific vertical locations of tonal peaks) does not simply vary gradiently across speakers and situations but constitutes a predictable part of the phonetic specification of tones.

4aSC31. Comparison of approaches to estimate the speech modulation transfer function. Karen L. Payton, Shaoyan Chen (ECE Dept., UMass Dartmouth, 285 Old Westport Rd., North Dartmouth, MA 02747-2300, kpayton@umassd.edu), and Louis D. Braida (MIT, Cambridge, MA 02139)

Using speech as a probe stimulus to compute the Speech Transmission Index (STI) has been of great interest to speech researchers. One technique is based on first computing the speech modulation transfer function (SMTF). Approaches used to obtain the SMTF include those developed by Steeneken and Houtgast [H. Steeneken and T. Houtgast, Proc. 11th ICA, Paris **7**, 85–88 (1983)] and Drullman *et al.* [R. Drullman, J. M. Festen, and R. Plomp, J. Acoust. Soc. Am. **95**, 2670–2680 (1994)]. This paper compares these two approaches and a new one to the theoretically obtained modulation transfer function (MTF) for reverberant and noisy environments. The new method computes the magnitude of the cross-power spectrum rather than the real part used by Drullman. As previously reported, Houtgast's method exhibits artifacts at high modulation frequencies. Drullman's approach eliminates artifacts in the reverberant environment but does not predict the theoretical MTF for the noisy environment. The new method outperforms the other two approaches in matching the theoretically derived MTF across both environments. This paper also examines the SMTF of amplitude-compressed speech for these three methods. [Work supported by NIDCD.]

4aSC32. The effect of variation in naturalness on phonetic perceptual identification. Robert E. Remez, Cynthia Y. Yang (Dept. of Psych., Barnard College, 3009 Broadway, New York, NY 10027-6598, remez@columbia.edu), Rebecca L. Piorkowski (Johns Hopkins Univ., Baltimore, MD 21218), Stephanie Wissig, Abigail Batchelder, and Heddy Nam (Barnard College, New York, NY 10027-6598)

The relation between apparent naturalness and phonetic identification was assessed in six perceptual tests. A seven-step place-of-articulation series spanning [da] to [ga] was created with speech synthesis approximating the spectra of natural samples. The sensitivity of perceivers to this realization of a place contrast was assessed by estimating the cumulative d' across the series in identification tests. Four variants of this series differing in apparent naturalness were produced by altering the synthesis source function while preserving the center frequency and bandwidth of the formants, and by replicating the gross spectrotemporal patterns with time-varying sinusoids. In addition to calibrating perceivers' sensitivity to the place contrast over variation in naturalness, we conducted a naturalness tournament composed of items drawn from the five test series. A correlation of the findings of the naturalness tournament with the measures of phonetic sensitivity offers an index of the effect of variation in naturalness on phonetic perception. This study can resolve the dispute between the classic premise that intelligibility and naturalness are orthogonal attributes of speech perception, and the more recent premise entailed by episodically based accounts of perceptual categorization, that novel instances are identified by virtue of auditory similarity to prior exemplars. [Research supported by NIDCD.]

4aSC33. Effects of bandpass noise and telephonic filtering on the perception of consonants. Thomas Purnell and Laura Kopplin (Dept. of Linguist., Univ. of Wisconsin, Madison, 1168 Van Hise, 1220 Linden Dr., Madison, WI 53706, tcpurnell@facstaff.wisc.edu)

Consonant identification studies report varying results of the usefulness of information above 4 kHz. A recent study found listeners misidentifying female-produced consonants characterized by a center of gravity above 3.5 kHz [P. G. Stelmachowicz *et al.*, J. Acoust. Soc. Am. **110**, 2183–2190 (2000)], suggesting that the perception of alveolar stops and fricatives as spoken by females are misperceived in common environments which filter out high-frequency spectral information. In the current study, consonant identification of female tokens consisting of consonants varying by place or manner ([p t k], [f T s S tS], [t tS s S n l r]) was tested in a forced choice experiment. Tokens were low-pass filtered at 3.5 kHz simulating telephonic filtering. Additionally, the locus of spectral information below 4 kHz was assessed by comparing responses to tokens masked by nonmodulating noise, where the signal-to-noise ratio was at least 10 dB SPL. Noise masking occurred between 0 and 0.4, 0.4 and 2, and 2 and 3.5 kHz. Confusion matrices indicated that the perception of unmasked speech is not hindered by low-pass filtering at 3.5 kHz, while perception of masked tokens was worse with masking between 2 and 3.5 kHz.

4aSC34. Consonant identification in noise: Effects of temporal asynchrony, noise fluctuations, and temporal envelope expansion. Frederic Apoux (LPE, UMR CNRS 8581, Inst. de Psychologie, Univ. Paris V, 71 Av E. Vaillant, 92774 Boulogne-Billancourt Cedex, France and INTRASON France S.A.), Stephane Garnier (Groupement d'audioprothsistes ENTENDRE-GIPA 2, 78760 Jouars-Pontchartrain, France), and Christian Lorenzi (Inst. de Psychologie, Univ. Paris V, 92774 Boulogne-Billancourt Cedex, France)

The effects of two temporal envelope expansion schemes on consonant identification performance were investigated with normal-hearing listeners. These effects were examined in the presence or absence of short-term energy fluctuations in the masker. The effects of adding a 500-ms forward fringe to the background noise was also investigated. Speech stimuli were 16 consonant–/a/ syllables. The expanded stimuli were obtained by raising their original envelope to the power k. In the first scheme, k was fixed at 2. In the second scheme, k varied from 4 to 0.3 as a function of the envelope power; the low-level parts of the envelope were therefore expanded whereas the high-level parts were compressed. As reported in many studies, identification was substantially improved by the presence of energy fluctuations in the masker. Adding the forward fringe generally degraded identification in fluctuating noise, but could improve identification in the stationary noise. Overall, expansion led to (i) poorer consonant identification when k was a constant and (ii) better consonant identification when k was varied. These effects were larger in stationary noise. This study suggests that consonant identification can be improved by a dual expansion/compression scheme. Some results obtained with hearing-impaired listeners will be presented.

4aSC35. Recognition of word and nonsense syllables in noise. José R. Benkí (Dept. of Linguist., Univ. of Michigan, Ann Arbor, MI 48109-1285)

In the recognition of human speech, listeners use sensory information from the speech signal as well as signal-independent information, such that acoustic-phonetic salience, lexical status, frequency of usage, and neighborhood density interact in speech recognition. This paper presents detailed results from a replication of a study of context effects reported by Boothroyd and Nittrouer [J. Acoust. Soc. Am. **84**, 101–108 (1988)]. 240 phonetically matched word and nonsense CVC syllables at different SNRs were presented to 32 listeners for identification. The results are consistent with the original study, with greater accuracy for words than nonsense items as quantified by a j-factor analysis according to lexical status. Online response collection enabled detailed analyses not reported in the original study. Values for the j-factor according to usage frequency and phonetic confusability are presented. Confusion matrices of the phonemes are presented, and are largely consistent with previous studies for the initial consonants and vowels. Accuracy in the final consonants of nonsense syllables is substantially lower than the corresponding phonemes of words or than the initial consonants of nonsense syllables. Final nasals were particularly difficult to identify.

4aSC36. Signal detection analysis of response distributions for intensity and speech judgments. Richard E. Pastore (SUNY at Binghamton, Binghamton, NY 13902-6000) and Neil A. Macmillan (Dept. of Psych., Brooklyn College (CUNY), Brooklyn, NY 11210-2889)

Schouten and van Hessen [J. Acoust. Soc. Am. **104**, 2980–2990 (1998)] reported rating response distributions for intensity and speech continua. These empirical distributions were claimed to reflect each subject's internal decision axes, allowing direct evaluation of the distribution assumptions underlying the common Gaussian equal-variance model signal detection theory (SDT). The observed distributions were deemed roughly Gaussian. However, the original analyses implicitly assume that the rating scale is a direct, interval reflection of internal likelihood values. We used the data instead to construct ROC curves, an analysis that assumes an ordinal relationship between the rating responses and the decision axes. Our ROC analyses show that the representations of the intensity stimuli reasonably approximate a Gaussian model. In contrast, the distributions underlying the speech stimuli are clearly not Gaussian, but imply discrete representations. We discuss appropriate decision models for judgments of intensity and speech continua.

4aSC37. Speech perception with tactile support in adverse listening conditions. Rob Drullman and Adelbert W. Bronkhorst (TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands)

Since long, different methods of vibrotactile stimulation have been used as an aid for speech perception by some people with severe hearing impairment. The fact that experiments have shown (limited) benefits proves that tactile information can indeed give some support. In our research program on multimodal interfaces, we wondered if normal hearing listeners could benefit from tactile information when speech was presented

in adverse listening conditions. Therefore, we set up a pilot experiment with a male speaker against a background of one, two, four or eight competing male speakers or speech noise. Sound was presented diotically to the subjects and the speech-reception threshold (SRT) for short sentences was measured. The temporal envelope (0–30 Hz) of the speech signal was computed in real time and led to the tactile transducer (MiniVib), which was fixed to the index finger. First results show a significant drop in SRT of about 3 dB when using tactile stimulation in the condition of one competing speaker. In the other conditions no significant effects were found, but there is a trend of a decrease of the SRT when tactile information is given. We will discuss the results of further experiments.

4aSC38. Audiovisual integration in the absence of a McGurk effect. Lawrence Brancazio (Dept. of Psych., Southern Connecticut State Univ., 501 Crescent St., New Haven, CT 06515 and Haskins Labs., New Haven, CT 06511), Joanne L. Miller, and Michèle Mondini (Northeastern Univ., Boston, MA 02115)

The McGurk effect, a change in perceived place of articulation due to an incongruent visual stimulus (e.g., auditory /pi/ with visual /ti/ perceived as /ti/), demonstrates the contribution of vision to speech perception. Interestingly, in a given experiment the McGurk effect typically does not occur on every trial. We investigated whether non-McGurk trials result from a failure to perceptually integrate auditory and visual information by simultaneously manipulating visual place of articulation and visual speaking rate. Previous work [Green and Miller, Percept. Psychophys. **38**, 269–276 (1985)] has shown that the boundary along an auditory /bi/–/pi/ voice-onset-time (VOT) continuum occurs at a longer VOT when the auditory stimulus is paired with a slow rather than a fast visual /pi/. We paired stimuli from an auditory /bi/–/pi/ continuum with fast and slow versions of a visual /ti/, and subjects identified each item as /b/, /p/, /d/, or /t/. We found a rate effect on McGurk trials, with the /d/–/t/ boundary occurring at a longer VOT when the visual stimulus was slow rather than fast. Importantly, we found a comparable rate effect for the /b/–/p/ boundary on non-McGurk trials. This indicates that audiovisual integration occurs even in the absence of a McGurk effect. [Work supported by NIH/NIDCD.]

4aSC39. Elderly perception of speech from a computer. Alan Black, Maxine Eskenazi, and Reid Simmons (Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213)

An aging population still needs to access information, such as bus schedules. It is evident that they will be doing so using computers and especially interfaces using speech input and output. This is a preliminary study to the use of synthetic speech for the elderly. In it twenty persons between the ages of 60 and 80 were asked to listen to speech emitted by a robot (CMU's VIKIA) and to write down what they heard. All of the speech was natural prerecorded speech (not synthetic) read by one female speaker. There were four listening conditions: (a) only speech emitted, (b) robot moves before emitting speech, (c) face has lip movement during speech, (d) both (b) and (c). There were very few errors for conditions (b), (c), and (d), but errors existed for condition (a). The presentation will discuss experimental conditions, show actual figures and try to draw conclusions for speech communication between computers and the elderly.

4aSC40. Virtual glides and duplex perception. Robert Allen Fox, Lawrence L. Feth, Ewa Jacewicz (Dept. of Speech and Hearing Sci., The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210-1002, fox.2@osu.edu), and Nandini Iyer (Univ. of Illinois, Champaign, IL 61820)

Duplex perception occurs when a tone glide is presented to one ear and an ambiguous synthetic CV is presented to the other ear. Listeners simultaneously perceive a syllable (e.g., /dɑ/ or /gɑ/) and a tone chirp. The present study examines whether duplex percepts can be generated when the tone glide presented is actually a virtual glide in which the frequency transition is evoked by modifying the amplitudes of two tones whose frequencies do not change [see Jacewicz *et al.*, J. Acoust. Soc. Am. **110**, 2657 (2002)]. A base synthetic CV token 250 ms in duration (with initial $F1$ and $F2$ transitions appropriate to both /d/ and /g/) was combined with a set of $F3$ transitions (representing either a synthetic formant transition, a true tone glissando, or a virtual glide transition) which produced a set of /dɑ/–/gɑ/ continua. These $F3$ transitions were either combined with the base and presented to both ears (diotic presentation) or presented simultaneously to different ears (dichotic presentation). Listeners heard the stimulus sets and were required to identify each token as /dɑ/ or /gɑ/. Preliminary results indicate that similar identification functions for all versions of $F3$ are generated in the monaural condition, but that these functions are shallower in the dichotic condition.

4aSC41. Relative roles of consonants and vowels in perceiving phonetic versus talker cues in speech. Gina Cardillo and Michael J. Owren (Dept. of Psych., Cornell Univ., Ithaca, NY 14853)

Perceptual experiments tested whether consonants and vowels differentially contribute to phonetic versus indexical cueing in speech. In 2 experiments, 62 total participants each heard 128 American–English word pairs recorded by 8 male and 8 female talkers. Half the pairs were synonyms, while half were nonsynonyms. Further, half the pairs were words from the same talker, and half from different, same-sex talkers. The first word heard was unaltered, while the second was edited by setting either all vowels (''Consonants-Only'') or all consonants (''Vowels-Only'') to silence. Each participant responded to half Consonants-Only and half Vowels-Only trials, always hearing the unaltered word once and the edited word twice. In experiment 1, participants judged whether the two words had the same or different meanings. Participants in experiment 2 indicated whether the word pairs were from the same or different talkers. Performance was measured as latencies and d values, and indicated significantly greater sensitivity to phonetic content when consonants rather than vowels were heard, but the converse when talker identity was judged. These outcomes suggest important functional differences in the roles played by consonants and vowels in normative speech.

4aSC42. Models of vowel perception. Michelle R. Molis[a] (Army Audiol. & Speech Ctr., Walter Reed Army Medical Ctr., Washington, DC 20307-5001)

The debate continues regarding the efficacy of formant-based versus whole spectrum models of vowel perception. Categorization data were obtained for a set of synthetic steady-state vowels and were used to evaluate both types of models. The models tested included various combinations of formant frequencies and amplitudes, principal components derived from critical-band spectra, and perceptually scaled LPC cepstral coefficients. The stimuli were 54 five-formant synthesized vowels that varied orthogonally in $F2$ (1081–2120 Hz) and $F3$ (1268–2783 Hz) frequency in equal 0.8 Bark steps. Fundamental frequency contour, $F1$ (455 Hz), $F4$ (3250 Hz), $F5$ (3700 Hz), and duration (225 ms) were held constant across all stimuli. Twelve speakers of American English (Central Texas dialect) categorized the stimuli as the vowels /I/, /U/, or /ɛ/. Results indicate that formant frequencies provided the best account of the data only when nonlinear terms were also included in the analysis. The critical-band principal components also performed reasonably well. While the principle of parsimony would suggest that formant frequencies offer the most appropriate description of vowels, the relative success of a richer,

more flexible and more neurophysiologically implementable model still must be taken into consideration. [Work supported by NIDCD.]
[a] Formerly at Univ. of Texas at Austin.

4aSC43. Patterns in the perception of VC(C)V strings. Terrance M. Nearey (Dept. of Linguist., Univ. of Alberta, Edmonton, AB T6G 2E7, Canada, t.nearey@ualberta.ca) and Roel Smits (Max Planck Inst. for Psycholinguist.)

A set of 144 /aC(C)a/ stimuli were constructed (inspired by, e.g., B. Repp, Percept. Psychophys. **24**, 471–485 (1978)]). Each consisted of: a vocoid (135 ms steady state followed by 50 ms VC closing transitions); a silent portion; and another vocoid (50 ms CV opening transition plus 250 ms steady state). Four different silent periods [80, 120, 190, and 300 ms] were crossed with six closing transition patterns (ranging from b-like to d-like closures) and with six analogous opening transitions. Thirteen listeners classified each stimulus ten times. Polytomous logistic regression showed that singletons, heterorganic clusters, and geminates all have distinct duration weights. Contrary to expectations, results for place were remarkably simple. Weights for closing transitions show labial closures (/b, bd, bb/) to be widely separated from the coronal closures (/d, db, dd/), but are otherwise homogeneous. Weights for opening cues show a similar dichotomy of labial releases /b, db, bb/ versus coronal releases (/d, bd, dd/); here, however, the singleton /b/ differs slightly from the other labials. A related experiment in Dutch will be discussed and results from both experiments will be compared against proposals from the literature and our own models. [Work supported by SSHRC.].

4aSC44. The effect of monauralized binaural displays on speech intelligibility. Robert S. Bolia (Air Force Res. Lab. (AFRL/HECP), 2255 H St., Wright–Patterson AFB, OH 45433-7022, robert.bolia@wpafb.af.mil)

It has often been demonstrated in laboratory settings that the spatial separation of multiple simultaneous talkers leads to enhanced intelligibility of a target message as well as an increase in the ease with which nontarget messages can be successfully ignored. This finding can be exploited in binaural intercom systems to improve communications' effectiveness. In some situations, operators use monaural displays at least part of the time, for example when operators in command and control nodes remove one earcup to talk to other nearby operators, in an effort to avoid squandering channels. The present investigation was designed to determine whether such a temporary monauralization of a binaural intercom would result in reduced speech intelligibility relative to a diotic control condition. Participants listened to 2 or 3 simultaneous presentations of phrases from the coordinate response measure (Bolia, Nelson, Ericson, and Simpson, 2000) and responded according to the instructions spoken by the target talker. The dependent measure was the percentage of correct responses. Independent variables included listening condition (monaural left, monaural right, binaural) and the locations of the target and distractor talkers. Results indicate that the performance in the monaural conditions were worse than in the binaural conditions but better than the diotic control.

4aSC45. Cerebral responses to across- and within-category change of vowel durations measured by near-infrared spectroscopy. Yasuyo Minagawa-Kawai (Independent Administrative Inst., Natl. Inst. of Japanese Lang.), Koichi Mori, Izumi Furuya, Ryoko Hayashi, and Yutaka Sato (Res. Inst., Natl. Rehabilitation Ctr. for the Disabled)

The present study examined cerebral responses to phoneme categories, using near-infrared spectroscopy (NIRS) by measuring the concentration and oxygenation of hemoglobin accompanying local brain activities. Targeted phonemes used here are Japanese long and short vowel categories realized only by durational differences. Results of NIRS and behavioral test revealed NIRS could capture phoneme-specific information. The left side of the auditory area showed large hemodynamic changes only for contrasting stimuli between which phonemic boundary was estimated (across-category condition), but not for stimuli differing by an equal duration but belonging to the same phoneme category (within-category condition). Left dominance in phoneme processing was also confirmed for the across-category stimuli. These findings indicate that the Japanese vowel contrast based only on duration is dealt with in the same language-dominant hemisphere as the other phonemic categories as studied with MEG and PET, and that the cortical activities related to its processing can be detected with NIRS. [Work supported by Japan Society for Promotion of Science (No. 8484) and a grant from Ministry of Health and Welfare of Japan.]

4aSC46. Importance of onset properties on spectral contrast for speech and other complex spectra. Jeffry A. Coady and Keith R. Kluender (Dept. of Psych., Univ. of Wisconsin, Madison, Madison, WI 53706, jacoady@facstaff.wisc.edu)

Previous studies using speech and nonspeech analogs have shown that auditory mechanisms that enhance spectral contrast serve to compensate for assimilation of spectral composition caused by coarticulated production. In the present studies, series of CV syllables varying acoustically in $F2$-onset frequency and perceptually from /ba/ to /da/ were identified either following [e] (higher $F2$) and [o] (lower $F2$) or following spectral complements of these vowels. Spectral complements had troughs replacing peaks corresponding to formants in the vowels. Both preceding vowels and spectral complements altered perception of the following stop, with complementary spectra providing results complementary to those for vowel spectra. However, effects for preceding spectral complements differed from those for preceding vowels by being dependent upon the precursor duration or inter-stimulus interval (ISI). Additional experiments using short spectral samples revealed that the single most important predictor of effects of vowel complements was the interval between the onset of the preceding sound and the onset of the following CV. This finding appears to implicate spectral characteristics of onsets (in addition to or instead of simple adaptation or masking) as being important for spectral contrast. The significance of these findings for the perception of connected speech will be described. [Work supported by NIDCD.]

4aSC47. Center-of-gravity effects in the perception of high front vowels. Ewa Jacewicz and Lawrence L. Feth (Dept. of Speech and Hearing Sci., The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210, jacewicz.1@osu.edu)

When two formant peaks are close in frequency, changing their amplitude ratio can shift the perceived vowel quality. This center-of-gravity effect (COG) was studied particularly in back vowels whose $F1$ and $F2$ are close in frequency. Chistovich and Lublinskaja (1979) show that the effect occurs when the frequency separation between the formants does not exceed 3.5 bark. The COG and critical distance effects were manifested when a two-formant reference signal was matched by a single-formant target of variable frequency. This study investigates whether the COG effect extends to closely spaced higher formants as in English /i/ and /I/. In /i/, the frequency separation between $F2$, $F3$, and $F4$ does not exceed 3.5 bark, suggesting the existence of one COG which may affect all three closely spaced formants ($F2=2030$, $F3=2970$, $F4=3400$ Hz). In /I/, each of the $F2$-$F3$ and $F3$-$F4$ separations is less than 3.5 bark but the $F2$-$F4$ separation exceeds the critical distance, indicating two COGs ($F2=1780$, $F3=2578$, $F4=3400$ Hz). We examine the COG effects using matching of four-formant reference signals, in which we change the

amplitude ratios, by two-formant targets with variable frequency of $F2$. The double-staircase adaptive procedure is used. [Work supported by an INRS award from NIH to R. Fox.]

4aSC48. Speaking rate adjustment across changes in talker. Rochelle S. Newman (Dept. of Speech & Hearing Sci., Univ. of Maryland, 0100 Lefrak Hall, College Park, MD 20742) and James R. Sawusch (Univ. at Buffalo, Buffalo, NY 14260)

Individuals vary their speaking rate, and listeners use the speaking rate of precursor sentences to adjust for these changes [Kidd (1989)]. Recent work has suggested that speaking rate adjustment may not always be limited to speech from a single talker (Sawusch and Newman, 2000). When a talker change occurs in the midst of a vowel, the durations of both portions of the vowel influence perception of the rate-based contrast. In the present experiments, we examined the effect of talker change on rate normalization for precursor phrases. A male talker produced the sentence, I heard him say the word—at one of three different speaking rates. A female talker then produced a nonword containing a duration-based contrast. We examined whether the male talker's speaking rate would influence perception of the female talker's speech. The results were somewhat surprising. The speaking rate of the first talker did influence perception of the contrast in the second talker. However, the effect was a U-shaped function of speaking rate, rather than the linear function typically demonstrated in the single-voice condition. Several follow-up studies replicated this basic pattern. Implications of this finding for studies of rate normalization will be discussed.

4aSC49. The effect of familiarity with a voice on continuous shadowing: Effects of explicit knowledge, but not of implicit familiarity. Rochelle S. Newman (Dept. of Hearing & Speech Sci. & Ctr. for Neurosci. & Cognit. Sci., Univ. of Maryland, College Park, MD 20742)

Listeners were asked to shadow a target voice while a second voice spoke fluently in the background. Listeners differed in their familiarity with the target voice: members of the first group of listeners were familiar with the voice and were told explicitly whose voice they would be hearing; members of a second group were familiar with the target voice but were not warned whose voice it was, and members of a third group were entirely unfamiliar with the target voice. Familiarity was from real-world experience: the voice was that of one of the participant's professors. Explicit knowledge of talker identity appeared to have the larger effect on listener performance: Participants in the first group made significantly fewer shadowing errors than those in the second group. In contrast, simple familiarity with a voice had no such influence. Familiarity did influence the types of errors listeners made, however: those participants who were familiar with the target voice prior to the test session made fewer incorrect responses than did those who had not heard the talker previously, although their total number of errors (including misses, as well as incorrect responses) did not differ.

4aSC50. Optimization of a three-dimensional Simplex procedure for speech in noise. Bas A. M. Franck, Jan Koopman, and Wouter A. Dreschler (Academic Medical Ctr., Dept. of Clinical and Exp. Audiol., D2-223, P.O. Box 1100 DD, Amsterdam, The Netherlands)

In this study we investigated means to enhance the reliability of a three-dimensional Simplex optimization procedure. The algorithms used to span the three dimensions are noise reduction, spectral tilting, and spectral enhancement, respectively. We considered the influence of an adaptive step size, different starting points, different types of noise, and the effect of an impaired spectral resolution. The subjects were asked to evaluate two consecutive, differently processed sentences on the evaluation criterion listening comfort. Three different noise types were used. We asked the normal-hearing subjects to listen to the stimuli with or without simulated impaired auditory filtering. The impairment was simulated by application of spectral smearing. For the subjects with normal filtering we applied a signal-to-noise ratio (S/N) of 0 dB. For the subjects with simulated impaired frequency resolution we applied S/N=5 dB. Due to several methodological innovations (especially when the step size is adaptive), the reliability of the Simplex procedure increased. Moreover, the reliability depends on the algorithms included and on the type of noise used.

4aSC51. Talker identification from analysis of raw complex waveforms. Michael A. Stokes (MAS Enterprises, 3929 Graceland Ave., Indianapolis, IN 46208, masmodel@indy.net)

Stokes (1996) demonstrated that visual inspection of raw complex waveforms can be used to identify a vowel produced by a talker. This research resulted in the MAS Model of Vowel Perception and Production (Stokes, 1998; http://home.indy.net/~masmodel/). More recently, another experiment extended this work to female talkers as well as male talkers (Stokes, 2001). Together, this research represents the only ongoing comprehensive research involving visual inspection of raw complex waveforms for identifying vowels produced by any talker. As an extension of the work, the present study involves identifying a talker from a waveform display. Unique voice signatures identified from waveform displays are used to identify a talker from a set of 10 talkers in the same way as one would identify a person from fingerprints. In two trials (the word who'd in trial 1 and heed in trial 2), a talker was correctly identified from a set of 10 unique talkers per trial using small visual samples of waveforms and matching it to a waveform sample of the talkers to be identified.

4aSC52. Mapping the perceptual space defined by transition onset frequencies of synthetic CV stimuli. John W. Hawks (Kent State Univ., Kent, OH 44242, jhawks@kent.edu), Emily Wead, and Marios Fourakis (Ohio State Univ., Columbus, OH 43210)

The perception of formant transition cues to place of articulation was investigated through the use of synthetic stop-vowel stimuli. Three sets of stop-vowel stimuli were synthesized. In each set, the $F2$ and $F3$ onset frequencies were varied in an orthogonal design resulting in 110 combinations. The sets differed in the $F2$ and $F3$ transition offset frequencies which were set to be appropriate for the vowels [i, u, a]. Twenty normal-hearing listeners identified each stimulus as beginning with [b], [d] or [g]. Listeners were also asked to rate the certainty of their responses on a scale of 1 (least certain) to 5 (most certain). While resembling stop-vowel sequences, no burst or aspiration parameters were included, thus forcing listeners to use only transition information as the primary cue to place of articulation. Perceptual spaces were constructed as defined by the $F2$ and $F3$ onset frequencies. In each perceptual space each $F2/F3$ onset combination was assigned the phonetic code [b], [d], or [g], reflecting the plurality of subjects' responses with certainty ratings higher than 3. The resulting spaces will be discussed in terms of the effect of the target vowel on the perception of $F2$ and $F3$ transitions.

Session 4aUWa

Underwater Acoustics: Reverberation and Scattering

Raymond Lim, Chair
Coastal Systems Station, 6703 West Highway 98, Panama City, Florida 32407-7001

Chair's Introduction—7:40

Contributed Papers

7:45

4aUWa1. Passive sonar calibration spheres. David M. Deveau (Naval Undersea Warfare Ctr. Det. AUTEC, 801 Clematis St., West Palm Beach, FL 33401)

The need for calibrated sonar targets is addressed with the development and testing of a set of thin-walled spheres filled with a high-density fluid. Using historical research information as a guide, a set of thin-walled metal spheres was developed and filled with a high-density fluid. The combination of the spherical shape and the acoustic focusing effects of the fluid enhanced the acoustic scattering strength of the shape so that it was not only stable with temperature but also significantly greater in amplitude. The simple passive nature of the spheres makes them ideal acoustic targets for long-term unattended deployments. Using a reference level measurement, each sphere was calibrated over a wide frequency range in order to provide the user with a curve of measured scattering strength versus frequency. The resulting curves showed a high degree of correlation between the individual spheres and the modeling that was used to extrapolate the theoretical values.

8:00

4aUWa2. Scattering of plane evanescent waves by cylindrical shells and wave vector coupling conditions for exciting flexural waves. Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

The coupling of sound to buried targets can be associated with acoustic evanescent waves when the sea bottom is smooth. To understand the excitation of flexural waves on buried shells by acoustic evanescent waves, the partial wave series for the scattering is found for cylindrical shells at normal incidence in an unbounded medium. The formulation uses the simplifications of thin-shell dynamics. In the case of ordinary waves incident on a shell, a ray formulation is available to describe the coupling to subsonic flexural waves [P. L. Marston and N. H. Sun, J. Acoust. Soc. Am. **97**, 777–783 (1995)]. When the incident wave is evanescent, the distance between propagating plane wavefronts is smaller than the ordinary acoustical wavelength at the same frequency and the coupling condition for the excitation of flexural waves on shells or plates is modified. Instead of matching the flexural wave number with the propagating part of the acoustic wave number only at the coincidence frequency, a second low-frequency wave number matching condition is found for highly evanescent waves. Numerical evaluation of the modified partial-wave-series appropriate for an evanescent wave is used to investigate the low-frequency coupling of evanescent waves with flexural wave resonances of shells.

8:15

4aUWa3. Low-frequency propagation and scattering from proud targets. Gary Sammelmann (Code R22 Coastal Systems Station, 6703 W. Hwy. 98, Panama City, FL 32407-7001)

This discussion focuses on the effects of placing a large aspect ratio target in the vicinity of an interface. It uses the spheroidal T-matrix description of scattering from a large aspect ratio target developed by Hackman to describe the scattering from large aspect ratio solids. A comparison of the effects of the interface on the form function and the directivity pattern of the scattering from the target at low frequencies is made. At sufficiently low frequencies the scattering from the target and its image add coherently, giving rise to a 6-dB increase in the form function of the proud target versus a volume target. At higher frequencies the scattering oscillates between constructive and destructive interface of the scattering from the target and its image. [Work supported by Codes 321OE and 32MIW of the ONR.]

8:30

4aUWa4. Scattering by a buried elongated object. Raymond Lim (Coastal Systems Station, Code R22, 6703 W. Hwy. 98, Panama City, FL 32407, limra@ncsc.navy.mil)

The feasibility of combining spherical and spheroidal basis sets in formulating a transition matrix for scattering by elongated objects embedded in an attenuating medium such as an ocean sediment is investigated. While a description completely in terms of spheroidal bases would likely yield optimal stability, spheroidal functions are more difficult to compute and well-tested algorithms are scarce for complex arguments. Therefore, a hybrid approach that expands the surface fields of the scatterer in spheroidal bases and all other exterior fields in spherical functions is tried. This approach mitigates the need for spheroidal functions of a complex argument. The merits of the hybrid approach are assessed by comparison against standard transition matrix formulations. [Work supported by ONR.]

8:45

4aUWa5. Low-frequency propagation and scattering in the ocean. Gary Sammelmann (Code R22 Coastal Systems Station, 6703 W. Hwy. 98, Panama City, FL 32407-7001)

This discussion describes various methodologies used to model the performance of low-frequency active systems with an emphasis on developing models for low-frequency signal-to-noise estimates and coherent imaging simulations. The global matrix method of Schmidt is adapted to define a numerically stable routine for finding the solutions of the characteristic equation for the normal modes. The effects of attenuation, surface

loss, and bottom loss are added through perturbation theory. Techniques for transforming the normal-mode equations into a time-domain representation are presented. Expansions of the normal-mode functions in terms of spherical and spheroidal basis functions at the source and receiver locations are discussed. A prescription for computing reverberation is presented, and a comparison between low-frequency and high-frequency models is presented. [Work supported by Codes 321OE and 32MIW of the ONR.]

9:00

4aUWa6. Broadband ocean reverberation coherence. Jon Reeves and Gary Morella (Appl. Res. Lab., Penn State Univ., P.O. Box 30, N. Atherton St., State College, PA 16804-0030)

Utilizing a wide-band transmitting/receiving system, at-sea experimental results for broadband reverberation have been acquired for both a harsh as well as a moderate ocean acoustic environment. Both stationary and moving sonar systems were employed in the study. Replica correlation (so-called match filter) processing of linear FM (LFM) pulse codes is employed to study the degree of reverberation coherence achievable over 4 octaves of bandwidth. The LFM signals are systematically processed over a bandwidth interval from 0.5 kHz to over 12 kHz.The corresponding reverberation coherence is displayed in terms of the variance of the cross correlation coefficient as a function of time-bandwidth product. Comparison with the theoretically expected variance for band-limited white Gaussian noise is also included. [Work sponsored by L. Jacobi, ONR 333.]

9:15

4aUWa7. Solving the background mystery in acoustical resonance scattering theory (RST). Changzheng Huang (Alfred Mann Inst., Univ. of Southern California, Los Angeles, CA 90089, changzh@bmsrs.usc.edu)

The acoustical resonance scattering theory (RST) aims to solving inverse scattering problems by decomposing the backscattering echoes into a background part and a resonance part. The former contains scatterers shape information, and the latter contains the material composition information. RST assumes the existence of an intermediate background for general scatterer and surrounding combinations. However, despite the efforts taken in the 1970s through 80s, such a background still remains elusive. This long-standing difficult problem is re-investigated in this study. It is concluded here that an intermediate background or alike is not needed for the purpose of solving inverse scattering problems. The simplest rigid or soft background is sufficient for the extraction of resonance information for any type of scatterer and surrounding combinations. In other words, the missing background problem is not a problem at all. Another new finding of this study is that a rational expression for the backscattering echoes is better suited to convey the resonance information than the additive decomposition expression as in RST. Explicit and concise relations exist between the rational expression and the acoustical properties of any scatterer, which is important for applications such as underwater target identification. This presentation will be illustrated by animation movies.

9:30

4aUWa8. The evolution of a pulse in a wave-guide and its interaction with a submerged structure: Some issues. Michael Werby (NRL, Code 7180, Stennis Space Center, MS 39529)

Initially let a pulse signal progress in time in a wave-guide along one direction *ab initio* at time t0. If it interacts with a submerged structure, at some time t1 then part of the wave will be refracted and part will be reflected so that at some point along the range between the object and the source there will be interference between the pulse signal and the reflected (backscattered) signal. This, to be sure, is a function of the duration of the pulse as well as other factors. Thus, in that region we must be required to deal with a two-way solution no matter what the topography simply because of the interference between the reflected wave and the time evolving wave. Here, the most general problem is properly formulated and the effect is examined as a function of pulse width, frequency (band) width, gating, and other factors. Results are presented in the form of frames or animations and it is observed that one may gain insight into this complicated process and it is possible to make use of this effect for extracting inverse information. [Work supported by ONR and NRL.]

9:45

4aUWa9. Rough surface reverberation of surface-generated ambient noise in a nonuniform ocean waveguide. Jin-Yuan Liu, Ping-Chang Hsueh, and Shern-Hsiung Tsai (Inst. of Undersea Technol., Natl. Sun Yet-sen Univ., Kaohsiung, Taiwan, ROC)

Rough surface reverberation of surface-generated ambient noise in an ocean waveguide with a transition layer of continuously varying density and sound speed is considered in this analysis. The general formulation is based upon a theory for rough surface scattering in a stratified medium, in conjunction with the analytical solutions for wave equation in a nonuniform medium intended to simulate the sediment layer in a realistic ocean. Numerical results for various parameters for the reverberation field were obtained and analyzed. The model is intended to refine the previously proposed one [J.-Y. Liu, H. Schmidt, and W. A. Kuperman, J. Acoust. Soc. Am. **93**, 753–769 (1993)], and will serve as a basis for analysis of the data obtained in a recent experiment. [Work supported by NSC of ROC.]

Session 4aUWb

Underwater Acoustics: Localization and Matched-Field Processing

Kevin D. Heaney, Chair
ORINCON Corporation, 4350 North Fairfax Drive, Suite 400, Arlington, Virginia 22203

Chair's Introduction—10:10

Contributed Papers

10:15

4aUWb1. Optimal source localization in a partially saturated fluctuating waveguide. Michele Zanolin and Nicholas C. Makris (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

When the propagation path between an acoustic source and receiving array is infested with natural disturbances, such as internal waves, part of the received field may be randomized by these waveguide fluctuations. When the fluctuations cause modal decoupling sufficient for this randomized portion of the field to obey the Central Limit Theorem, it will follow a circular complex Gaussian random process in time. The total field is then said to be partially saturated. In this case the signal is partially deterministic and partially random so that both the mean and covariance of the field depend on the source position. The asymptotic bias, covariance, and signal-to-noise ratio presented in Naftali and Makris [J. Acoust. Soc. Am. **110**, 1917–1930 (2001)] for a Gaussian random field are extended to treat this more general case. Analytic expressions and simulations are performed to determine the minumum signal-to-noise ratio necessary to optimally localize an acoustic source in practical scenarios in typical continental shelf environments as a function of saturation level. The specific scenarios investigated are for the New Jersey Strataform area which serves as a natural laboratory for numerous Office of Naval Research Ocean Acoustic Experiments.

10:30

4aUWb2. Localization when the source function is known. Laurie T. Fialkowski, Joseph F. Lingevitch, Gregory J. Orris, Michael Nicholas (Naval Res. Lab., Washington, DC 20375, lfialkowski@milton.nrl.navy.mil), and John S. Perkins (Naval Res. Lab., Washington, DC 20375)

Frazer and Pecholcs [J. Acoust. Soc. Am. **88**, 995–1002 (1990)] have investigated single-hydrophone time-domain localization. They introduced several localizers, similar to an earlier method introduced by Clay [J. Acoust. Soc. Am. **81**, 660–664 (1987)], that differ in the amount of *a priori* information required about the source function. In this paper we revisit these localizers to look at the special case where the source function is known and explore the benefits of multiple hydrophones. This case is of interest due to its similarity to active matched field processing and because these localizers can be used as cost functions in environmental inverse problems when the source position is known precisely. We apply the localizers to data that includes several different types of source waveforms, and was collected in shallow water southwest of Key West, Florida. [Work supported by ONR.]

10:45

4aUWb3. Passive phase-conjugate signaling for autonomous systems in the ocean. Paul Hursky, Martin Siderius, Michael B. Porter (SAIC, 1299 Prospect St., Ste. 305, La Jolla, CA 92037), Vincent K. McDonald, and Joseph A. Rice (SPAWAR Systems Ctr., San Diego, CA 92152-5001)

Acoustic communications is an attractive way to wirelessly network distributed assets, such as UUVs, autonomous sensors, and other systems in the ocean. We have been experimenting with modulation techniques based on passive phase-conjugation (PPC), which essentially uses the channel itself to refocus multipath arrivals, that would otherwise cause inter-symbol interference. We have also used Gold codes to isolate snapshots of the channel response, prior to the PPC operation. Our scheme can be used point-to-point, without arrays. This is a very simple demodulation technique which nevertheless is coherent and can overcome signal distortion due to multipath. Simplicity is very important, because simple algorithms consume less battery power, allowing the node to operate for a longer time in the water. We discuss details of our algorithms, including synchronization and Doppler compensation, and present the results of testing them in a variety of environments, including four SignalEx tests and the ASCOT01 experiment. We compare our PPC signaling scheme to frequency-hopped frequency-shift keying and direct sequence spread spectrum modulation using a rake receiver. All of these signaling schemes are being considered for low power, lightweight applications.

11:00

4aUWb4. Matched field processing in a highly variable shallow water site. Martin Siderius, Michael B. Porter, Paul Hursky (SAIC, 1299 Prospect St., La Jolla, CA 92037), Peter L. Nielsen (SACLANT Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy), and Jurgen Sellschopp (Forschungsanstalt fur Wassershall- und Geophysik, Klausdorfer Weg 2-24, 24148 Kiel, Germany)

Variability in the ocean environment can have a big impact on acoustic propagation. Acoustic receptions often contain multi-path contributions with fluctuations that vary significantly from the direct path to the higher order multi-path. Matched-field processing (MFP) is a beamforming method that takes advantage of the multi-path to extract information about the location of the sound source. Matched-field processing is generally successful in environments that are well known and can be incorporated into propagation models. However, in some cases the environmental conditions are too extreme for good propagation predictions. To explore the limits of MFP in highly variable environments, an experiment (ASCOT-01) was conducted last June off the New England coast. The acoustic measurements were made over several days between a sound source and a moored vertical line array of receivers. Extensive environmental measurements were made including the use of a vertical CTD chain (conductivity, temperature and depth) that was towed through the propagation path to provide a detailed view of the ocean sound speed profile. Results characterizing the difficulties of MFP at this site will be presented. Standard estimators will be compared to new alternatives intended to be robust against harsh environmental conditions.

11:15

4aUWb5. Time reversal, back propagation, matched field processors, correlation receivers, and the principles of radar/sonar signal design. Arthur Baggeroer (MIT, Rm. 5-204, Cambridge, MA 02139)

Time reversal and backpropagation have been demonstrated in several experiments. Similarly, while matched field processing (MFP) differs in terms of implentation—experimental vs computed replicas—both have

two common properties: (i) they are based on a single, spatially coherent signal; and (ii) the conjugate transpose of the Green's function and replica correlation are identical for self-adjoint systems. Hence, the principles for focusing and ambiguity plane properties of these processors are virtually identical to those for correlation receivers. The principles of optimal signal design for correlation receivers were the subject of much research for radar/sonar systems four decades ago and many of them seem to have been neglected in the analysis of time reversal, back propagation, and matched field processors. For example, time reversal from a point, a line array, or a random array of scatterers are duals of an impulse, a frequency modulated, and a pseudo-random noise signal, respectively. The equivalence and consequences of the time-bandwidth products for signals and array length wave number spread are demonstrated. The impact of sidelobes and multipath spread can be predicted. The generalizations of the important radar/sonar uncertainty principle, however, have yet been not demonstrated. This presentation reviews these optimal signal design principles and applies them to time reversal and MFP.

11:30

4aUWb6. Analytic prediction of sidelobe statistics for matched-field processing. Brian Tracey, Nigel Lee, and Lisa Zurk (Lincoln Lab., MIT, 244 Wood St., Lexington, MA 02420, btracey@ll.mit.edu)

Underwater source localization using matched-field processing (MFP) is complicated by the relatively high sidelobe levels characteristic of MFP ambiguity surfaces. An understanding of sidelobe statistics is expected to aid in designing robust detection and localization algorithms. MFP sidelobe levels are influenced by the underwater channel, array design, and mismatch between assumed and actual environmental parameters. In earlier work [J. Acoust. Soc. Am. **108**, 2645 (2000)], a statistical approach was used to derive analytic expressions for the probability distribution function of the Bartlett ambiguity surface. The distribution was shown to depend on the orthogonality of the mode shapes as sampled by the array. Extensions to a wider class of array geometries and to broadband processing will be shown. Numerical results demonstrating the accuracy of the analytic results and exploring their range of validity will be presented. Finally, analytic predictions will be compared to data from the Santa Barbara Channel experiment. [Work sponsored by DARPA under Air Force Contract F19628-00-C0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the Department of Defense.]

11:45

4aUWb7. Complex spatial envelopes, invariants and matched field processing. Kevin D. Heaney and Henry Cox (ORINCON Corp., 4350 N. Fairfax Dr., Ste. 400, Arlington, VA 22203)

The waveguide invariant was introduced in the Russian Literature (S. D. Chuprov, Ocean Acoustics, 1982) to explain range-frequency interference patterns of intensity. The concept of the spatial envelope is introduced here to show how the waveguide invariant can be derived and interpreted in terms of pressure rather than intensity. It is shown how this leads to range-frequency scaling that can be used to generate finely spaced (in frequency) matched field replicas from course spaced propagation calculations. For a downward refracting sound speed profile the acoustic propagating modes can be separated into groups (refracted bottom interacting, surface-bottom interacting). These mode groups have very similar waveguide invariants. The frequency scaling can be applied to these mode groups separately. An example is presented where 150 Hz shallow water replicas are computed every 20 Hz and interpolated, yielding a factor of 30 in reduction of replica computation for MFP.

12:00

4aUWb8. Nonspecular sound reflection and phase conjugation. Leonid M. Lyamshev (Andreev Acoust. Inst., Shvernik St. 4, Moscow 117036, Russia, lyamshev@kapella.gpi.ru)

A strong back reflection is observed at certain angles of incidence of sound waves at a limited plate or shell in a liquid. This reflection is termed nonspecular reflection. It arises whenever the condition for coincidence or spatial resonance in a plate is satisfied, i.e., in other words, when the velocity of the incident wave trace coincides with the propagation velocity of any free wave in a plate or shell. If the surface of a plate or shell performs forced vibrations at the frequency doubled in comparison with the frequency of the incident wave, back reflection is observed at any incidence angle. This phenomenon is termed phase conjugation. It is caused by nonlinear interaction of an incident wave with the vibrations of a plate or shell. Fundamentals of nonspecular reflection and phase conjugation are given, i.e., theoretical concepts and some experimental data. The generality and differences of the phenomena of nonspecular reflection and phase conjugation are discussed as well as their importance for practice. [Work supported by the Russian Foundation for Basic Research.]

Session 4pAA

Architectural Acoustics and Psychological and Physiological Acoustics: Perceptual Aspects of Real and Virtual Rooms

Barbara G. Shinn-Cunningham, Cochair
Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, Massachusetts 02215

Murray R. Hodgson, Cochair
Occupational Hygiene Program, University of British Columbia, 2206 East Mall, Vancouver, British Columbia V6T 1Z3, Canada

Chair's Introduction—1:30

Invited Papers

1:35

4pAA1. Challenges and solutions for realistic room simulation. Durand R. Begault (NASA Ames Res. Ctr., M.S. 262-2, Moffett Field, CA 94035, dbegault@mail.arc.nasa.gov)

Virtual room acoustic simulation (auralization) techniques have traditionally focused on answering questions related to speech intelligibility or musical quality, typically in large volumetric spaces. More recently, auralization techniques have been found to be important for the externalization of headphone-reproduced virtual acoustic images. Although externalization can be accomplished using a minimal simulation, data indicate that realistic auralizations need to be responsive to head motion cues for accurate localization. Computational demands increase when providing for the simulation of coupled spaces, small rooms lacking meaningful reverberant decays, or reflective surfaces in outdoor environments. Auditory threshold data for both early reflections and late reverberant energy levels indicate that much of the information captured in acoustical measurements is inaudible, minimizing the intensive computational requirements of real-time auralization systems. Results are presented for early reflection thresholds as a function of azimuth angle, arrival time, and sound-source type, and reverberation thresholds as a function of reverberation time and level within 250-Hz–2-kHz octave bands. Good agreement is found between data obtained in virtual room simulations and those obtained in real rooms, allowing a strategy for minimizing computational requirements of real-time auralization systems.

2:05

4pAA2. Effects of reverberation on spatial auditory performance and spatial auditory cues. Barbara Shinn-Cunningham (Hearing Res. Ctr., Depts. of Cognit. and Neural Systems and Biomed. Eng., Boston Univ., 677 Beacon St., Boston, MA 02215) and Norbert Kopčo (Boston Univ., Boston, MA 02215)

Ordinary room echoes and reverberation influence sound localization performance, moderately degrading judgments of source direction (i.e., judgments of azimuth and elevation) while vastly improving judgments of source distance. The relative magnitude of the influence of room acoustics on localization performance depends on the location of the listener in a room; specifically, performance is best when the listener is far from any wall (when there are no early, intense echoes). However, in addition to depending on listener location, localization performance in all three spatial dimensions depends on past experience in a particular room and improves over time. This room learning (which is not seen in a comparable anechoic study) at least partially generalizes across listener positions in a room. After reviewing results of localization studies demonstrating these effects, acoustic analyses are presented that show how echoes and reverberation alter and distort the spatial acoustic cues reaching the listener in an ordinary room and how these distortions vary with listener location in the room. Possible mechanisms to explain how spatial auditory processing might adapt in different acoustic environments in order to improve localization accuracy will be discussed. [Work supported in part by AFOSR Grant F49620-01-1-0005 and the Alfred P. Sloan Foundation.]

2:35

4pAA3. The direct-to-reverberant ratio as cue for distance perception in rooms. Adelbert W. Bronkhorst (TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands, bronkhorst@tm.tno.nl) and Pavel Zahorik (Waisman Ctr., Univ. of Wisconsin, Madison, WI 53705)

Although it is evident that in order to localize a sound source, one has to perceive both its direction and its distance, localization research has for many years focused mainly on the directional component. This situation has changed in the last few years. Recent studies have provided more insight into distance perception and the specific cues that are used by the auditory system. It appears that auditory and visual distance perception are similar, because both rely on a combination of cues, which are weighted differently in different environments. One auditory cue that has received particular attention is the direct-to-reverberant energy (D/R) ratio, because it is thought to be relatively stable, particularly in rooms. At present, however, it is not clear what the precise contribution of this cue

is or how it is coded in the auditory system. Evidence that the cue is derived from the monaural time-domain signal is contradicted by other data suggesting that it relies on binaural information. Furthermore, direct measurement of JNDs for the D/R ratio reveals unexpectedly poor performance. Merging these seemingly incompatible results into a coherent model is challenging, but necessary for obtaining a clearer picture of how the D/R cue contributes to perceived distance.

3:05

4pAA4. Investigation of reverberation synthesized by electro-acoustic enhancement systems, from a subjective and physical acoustic standpoint. Yasushi Shimizu (Architectural Acoust. Prog., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180)

Current electro-acoustic enhancement technology enables wide control over concert hall acoustics. The goal of sound field synthesis in anechoic space is to reconstruct a specific sound field. However, applying acoustic enhancement technology to existing reverberant spaces is a less developed research direction. This presentation demonstrates a methodology of electro-acoustic enhancement using regenerative reverberation through SAAF (spatially averaged acoustic feedback), an acceptable variation of RT and SPL in enhanced acoustical conditions. That was presented by YAMAHA Acoustic Research Laboratories. SAAF technology can flatten amplitude peaks at the howling frequency of acoustical feedback loops by using time variant finite impulse response filters. Therefore it enables regeneration of reverberated sound by wide band feedback in frequency without coloration. This system has been applied to ''negative absorption control'' and loudness equalization of under-balcony seats in current concert halls, to optimize concert hall acoustics electronically instead of architecturally. Adjusted reverberation time in enhanced condition should be between 1.5 and 2.0 times higher than the natural RT (ex. RTon/Rtoff=1.8). The SPL increases about 1 dB to 3 dB based on measured results of more than 30 performing halls integrated with acoustic enhancement system in Japan. Examples of the major Japanese concert halls with acoustic enhancement systems are presented.

3:35–3:50 Break

Contributed Papers

3:50

4pAA5. Reverberation and the Franssen effect. William M. Whitmer, William A. Yost, and Raymond H. Dye, Jr. (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626, wwhitme@luc.edu)

The Franssen effect, in which the location of a sudden-onset (transient) tone can occlude the location of a contralateral slow-onset (steady-state) tone, has been previously shown to occur only in a reverberant space. The nature of the reverberation required—how much and what kind—for the effect to occur has yet to be determined. To explore the role of reverberation in the Franssen effect, listeners were asked to identify the location(s) of transient/steady-state pure tones with frequencies ranging from 250–4000 Hz in a variety of real and virtual contexts: (a) limited (single reflective panel) reverberant enclosure; (b) discrete reflections in an anechoic room; (c) mannequin recordings of a reverberant enclosure presented over headphones; (d) impulse-response filtered signals presented over headphones; and (e) simulated 3-D reverberation via consumer-grade software. Results indicate that the Franssen effect can be a simple, effective means to judge the verisimilitude of a virtual environment. [Work supported by NIDCD.]

4:05

4pAA6. School cafeteria noise—The impact of room acoustics and speech intelligibility on children's voice levels. Joseph F. Bridger (Stewart Acoust. Consultants, P.O. Box 30461, Raleigh, NC 27622, joe@stewartacousticalconsultants.com)

The impact of room acoustics and speech intelligibility conditions of different school cafeterias on the voice levels of children is examined. Methods of evaluating cafeteria designs and predicting noise levels are discussed. Children are shown to modify their voice levels with changes in speech intelligibility like adults. Reverberation and signal to noise ratio are the important acoustical factors affecting speech intelligibility. Children have much more difficulty than adults in conditions where noise and reverberation are present. To evaluate the relationship of voice level and speech intelligibility, a database of real sound levels and room acoustics data was generated from measurements and data recorded during visits to a variety of existing cafeterias under different occupancy conditions. The effects of speech intelligibility and room acoustics on childrens voice levels are demonstrated. A new method is presented for predicting speech intelligibility conditions and resulting noise levels for the design of new cafeterias and renovation of existing facilities. Measurements are provided for an existing school cafeteria before and after new room acoustics treatments were added. This will be helpful for acousticians, architects, school systems, regulatory agencies, and Parent Teacher Associations to create less noisy cafeteria environments.

4:20

4pAA7. Comparison of two speech privacy measurements, articulation index (AI) and speech privacy noise isolation class (NIC′), in open workplaces. Heakyung C. Yoon and Vivian Loftness (School of Architecture, Carnegie Mellon Univ., Margaret Morrison Carnegie Hall, Rm. 410, Pittsburgh, PA 15213, hcy@andrew.cmu.edu)

Lack of speech privacy has been reported to be the main dissatisfaction among occupants in open workplaces, according to workplace surveys. Two speech privacy measurements, Articulation Index (AI), standardized by the American National Standards Institute in 1969, and Speech Privacy Noise Isolation Class (NIC′, Noise Isolation Class Prime), adapted from Noise Isolation Class (NIC) by U. S. General Services Administration (GSA) in 1979, have been claimed as objective tools to measure speech privacy in open offices. To evaluate which of them, normal privacy for AI or satisfied privacy for NIC′, is a better tool in terms of speech privacy in a dynamic open office environment, measurements were taken in the field. AIs and NIC's in the different partition heights and workplace configurations have been measured following ASTM E1130 (Standard Test Method for Objective Measurement of Speech Privacy in Open Offices Using Articulation Index) and GSA test PBS-C.1 (Method for the Direct Measurement of Speech-Privacy Potential (SPP) Based on Subjective Judgments) and PBS-C.2 (Public Building Service Standard Method of Test Method for the Sufficient Verification of Speech-Privacy Potential (SPP) Based on Objective Measurements Including Methods for the Rating of Functional Interzone Attenuation and NC-Background), respectively.

4:35

4pAA8. Speech intelligibility prediction of room with curved surfaces by means of its physical characteristics. Mokhtar Harun, Tharek Abdul Rahman, Md. Najib Ibrahim, and Md. Yunus Jaafar (Universiti Teknologi Malaysia, 81310 Skudai Johor, Malaysia)

In the room, it was found previously that the ear uses the first arriving sound to locate the source, but the early reflections contribute to the intelligibility. Late reflections, on the other hand reduce intelligibility. Therefore, for speech intelligibility a high sound energy fraction is necessary, requiring high early energy and low late energy. Spatial impression which is the subjective effect of early sound reflections that come from the side walls and ceiling, contributes to the sense of envelopment. This paper presents the investigation of speech intelligibility with respect to the physical characteristics of rooms with curved surfaces. The investigated physical characteristics are source-to-listener distance (Lr), listener-to-the farthest wall (Lwmax), listener-to-the closest wall (Lwmin), listener-to-the farthest wall and listener-to-the average distance to walls (Lwave). In addition to objective acoustics measures such as reverberation time (RT60), early decay time (EDT), average sound absorption coefficient (Aave) and clarity (C50), speech intelligibility predictors STI and Alcons were also measured and analyzed. It is the intention of this paper to explore the possibility of predicting speech intelligibility performance of a room with curved surfaces by using its physical characteristics.

4:50

4pAA9. Analysis of the uncertainty budget for the reflection and absorption coefficient given by the two-microphone method. Adolfo Esquivel D., Asoc and Guillermo Silva P. (Centro Nacional de Metrologia, km 4.5 Carr, a Los Cues, Mpio. El Marques, Queretaro 76241, Mexico)

The method for measurement of the reflection and absorption coefficient at normal incidence consists of a sound source at one end of an impedance tube, where the plane waves propagated are reflected from a sample that is at the other end of the tube. Those plane waves are generated by a random signal, and the decomposition of the standing wave pattern is carried out measuring the acoustical pressure at two-fixed microphone positions, from which it is possible to determine a transfer function to obtain the reflection and absorption coefficient. The systematic errors associated with the two-microphone technique using an impedance tube are analyzed, i.e., microphone-sample distance, microphone spacer, atmospheric measurement conditions, correction factors to the microphone sensitivity, and phase and amplitude mismatch in the microphones. These parameters constitute some of the sources to the measurement uncertainty of the measurand-reflection and absorption coefficient. The estimation of the measurement uncertainty for the reflection and absorption coefficient by the two-microphone method; likewise the analysis of the uncertainties sources and its sensitivity coefficients are discussed using a linear approach model. Furthermore, the influence of the correction factor due to the frequency response mismatch in the microphones in the calculation of the reflection and absorption coefficient is experimentally evaluated.

THURSDAY AFTERNOON, 6 JUNE 2002 LE BATEAU ROOM, 2:00 TO 4:50 P.M.

Session 4pBB

Biomedical Ultrasound/Bioresponse to Vibration: Characterization and Monitoring of Biological Flows/ Vibrational Analysis of Biological Systems

E. Carr Everbach, Cochair
Department of Engineering, Swarthmore College, 500 College Avenue, Swarthmore, Pennsylvania 19081-1397

Constantin C. Coussios, Cochair
Biomedical Engineering, University of Cincinnati, Medical Sciences Building, 231 Albert Sabir Way, Cincinnati, Ohio 45267-0761

Chair's Introduction—2:00

Invited Papers

2:05

4pBB1. Modeling strategies of ultrasound backscattering by blood. Cloutier Guy, Savery David, Fontaine Isabelle, and Teh Beng Ghee (Res. Ctr., Univ. of Montreal Hospital, 2099 Alexandre de Seve, Rm. Y-1619, Montreal, QC H2L 2W5, Canada)

Tissue characterization using ultrasound (US) scattering can allow the identification of relevant cellular biophysical information noninvasively. The characterization of the level of red blood cell (RBC) aggregation is one of the proposed applications. Different modeling strategies have been investigated by our group to better understand the mechanisms of US backscattering by blood, and to propose relevant measurable indices of aggregation. It could be hypothesized from these studies that the microstructure formed by RBC clusters is a main determinant of US backscattered power. The structure factor, which is related to the Fourier transform of the microscopic density function of RBCs, is described and used to explain the scattering behavior for different spatial arrangements of nonaggregated and aggregated RBCs. The microscopic density function was described by the Percus–Yevick approximation (nonaggregated RBCs), and for aggregated RBCs, by the Poisson distribution, the Neyman–Scott point process, and very recently by a flow-dependent rheological model. These statistical and microrheological models allowed the study of US backscattered power as a function of the hematocrit, scatterers' size, insonification frequency, and level of RBC aggregation. Experimental results available from the literature were used to validate the different approaches. [Work supported by Canadian Institutes of Health Research (MOP-36467), HSFQ, FCAR, and FRSQ.]

2:35

4pBB2. Ultrasonic scattering by blood: theories, experimental results and biomedical applications. K. Kirk Shung (Dept. of Bioengineering, 231 Hallowell Bldg., Penn State Univ., University Park, PA 16802, kksbio@engr.psu.edu)

In this paper, theoretical and experimental efforts that have been undertaken to better understand the phenomenon of ultrasonic scattering in blood will be reviewed. This subject is of interest in biology and medicine because the echoes generated by blood are used to extract blood velocity by ultrasonic Doppler flow and imaging devices. In the course of these investigations it became clear that ultrasonic scattering from blood is dependent upon such hematological and hemodynamic properties of blood as hematocrit, plasma protein concentration, flow rate and flow cycle duration. Several aspects of these experimental results have been successfully modeled by recent theoretical developments. An unexpected consequence of these efforts is that ultrasound appears to be a viable tool for blood flow visualization and hemodynamic measurements. Two unique hemodynamic phenomena that have never been reported in the hemodynamic literature have been observed: the black hole, a low echogenic zone in the center stream of whole blood flowing in a blood vessel under steady flow and the collapsing ring, an echogenic ring appearing near the periphery of a vessel at the beginning of a flow cycle, converging toward the center, and eventually collapsing during pulsatile flow. Similar observations have been made during clinical scanning of patients.

3:05

4pBB3. Blood velocity estimation by Doppler ultrasound: Problems and issues. Richard S. C. Cobbold and Aaron H. Steinman (Inst. of Biomaterials & Biomed. Eng., Univ. of Toronto, Toronto, ON M5S 3G9, Canada)

Modern methods of estimating blood flow have advanced considerably since Dr. Satamura and his colleagues at Osaka University reported the first measurements in 1959 using CW ultrasound. Most current methods involve the use of linear or phased arrays, which make possible 2-D color flow mapping, measurements from a single sample volume, and simultaneous B-mode imaging. However, along with this technology, troubling reports have been published (including our own) concerning the accuracy with which the velocity can be estimated and this has serious potential consequences in the quantitative assessment of vascular disease. The cause has not yet been completely identified, but appears to be partially associated with the wide range of Doppler angles within the sample volume, the complexity of the ultrasound propagation process, and the stochastic nature of the red blood cells flowing in the blood vessels. These causes will provide a focus for our review of the current status of Doppler ultrasound for vascular disease assessment and suggestions for future research. Specifically, a complete model of the entire measurement system is needed, and this includes the beamforming architecture, signal processing, possible nonlinear effects, scattering process, and nature of the 3-D vector flow field being measured.

3:35–3:50 Break

Contributed Papers

3:50

4pBB4. Measurement of volumetric flow rates with ultrasonic contrast agents using a depletion method. Jeffrey A. Ketterling, Cheri X. Deng, Frederic L. Lizzi (Riverside Res. Inst., 156 William St., New York, NY 10038), Ronald H. Silverman, and D. Jackson Coleman (Weill Medical College of Cornell Univ., New York, NY 10021)

A high-intensity focused transducer was used to deplete a small, well-defined volume of contrast agent flow in a small diameter tube. The depleted volume was probed both confocally and downstream with a 7.5-MHz focused diagnostic transducer. The depletion volume was found from theoretical calculations of the acoustic beam profile. Volume flow rates were then calculated by measuring the time required for the depleted volume to pass the diagnostic transducer. Additionally, a beam with known sidelobe spacing can give information about the flow velocity if the high-intensity transducer exposure is short and the sidelobes form spatially separated subdepletion regions. A 4.7-MHz ''strip electrode'' high-intensity transducer with 0.5-mm sidelobe spacing was used to study the flow velocity in dialysis tubing with a 3.1-mm outer diameter. Flow of known volume velocity was generated with a syringe pump. Both Optison and Albunex contrast agents were used. Experimental flow velocities were in excellent agreement with the known flow measurement.

4:05

4pBB5. Surface response of a viscoelastic medium to subsurface acoustic sources with application to medical diagnosis. Yigit Yazicioglu, Thomas J. Royston, and Francis Loth (Univ. of Illinois at Chicago, 842 W. Taylor St., MC 251, Chicago, IL 60607, troyston@uic.edu)

The response at the surface of an isotropic viscoelastic halfspace to buried fundamental acoustic sources is studied theoretically, computationally and experimentally. Finite and infinitesimal monopole and dipole sources are considered. Analytical and numerical integral solutions that account for compression, shear and surface wave response to the buried sources are formulated and compared with numerical finite element simulations and experimental studies on finite dimension phantom models. The focus here is on developing a better understanding of how biological soft tissue affects the transmission of vibro-acoustic energy from biological acoustic sources below the skin surface, such as a turbulent regime in an artery caused by a partial blockage. Such an understanding could catalyze the development of noninvasive procedures using an array of acoustic sensors on the skin surface for the identification and monitoring of vascular blockages a precursor to many serious cardiovascular diseases. [Research supported by NIH NCRR Grant #14250 and the Whitaker Foundation.]

4:20

4pBB6. Coupled vibration analysis of large arterial vessels. Xiaoming Zhang, Mostafa Fatemi, and James F. Greenleaf (Ultrasound Res. Lab., Mayo Clinic, Rochester, MN 55905, zhang.xiaoming@mayo.edu)

A coupled vibration model, which takes into account the elastic arterial wall, the interior blood, and exterior tissue, is presented. The arterial wall is considered as a cylindrical elastic shell and is modeled with the thick shell theory due to the large thickness-to-radius ratios. Both the blood and the tissue are considered as ideal fluids and are modeled with acoustic wave theory. A newly developed method, the wave propagation approach, is extended for the coupled vibration of arterial walls. The theoretic model is evaluated against those available in the literature. The experiments were carried out on a silicone rubber tube in a water tank with an ultrasound stimulated system. This system uses the radiation force of ultrasound to vibrate the tube at low frequency and records the resulting response by a

laser vibrometer. Both the excitation and measurement are remote and noncontact. The measured fundamental frequency is in good agreement with the theoretical prediction.

4:35

4pBB7. A USAE frequency shift method for monitoring stiffness variations in tissues. Elisa Konofagou, Kullervo Hynynen (Brigham and Women's Hospital, Harvard Med. School, Boston, MA 02115), and Jonathan Thierman (Harvard–MIT Div. of Health Sci. and Technol., Cambridge, MA)

In solid mechanics, the resonance frequency of a material is typically linked to its mechanical properties. A frequency shift therefore typically denotes a change in stiffness. It has already been shown that temperature variation causes stiffness changes in tissues. The amplitude of the ultrasound-stimulated acoustic emission (USAE) signal has also been shown to vary with temperature as well as indicate temperature elevation. Its amplitude, however, is sensitive to both acoustical and mechanical parameters, which at most frequencies have opposite effects resulting from temperature changes. In this paper, the measurement of a frequency shift of the USAE resonant peaks is used for monitoring of the tissue stiffness variation with temperature. In a numerical simulation, the variation of the frequency shift (ranging between -150 and 500 Hz) at different temperatures is shown. Then, in a series of experiments involving a gel phantom and porcine muscle tissue, the frequency shift variation (ranging between -200 and 200 Hz) is shown to follow the known stiffness changes due to temperature. Both simulation and experimental results indicate that the USAE frequency shift method can dissociate the mechanical from the acoustical parameter dependence and detect tissue coagulation. [Work supported by NIH Grant CA82275.]

THURSDAY AFTERNOON, 6 JUNE 2002 BRIGADE ROOM, 2:00 TO 4:45 P.M.

Session 4pEA

Engineering Acoustics: Transducers, Materials and Modeling

Harold C. Robinson, Chair
NAVSEA Division, Naval Undersea Warfare Center, Code 2131, 1176 Howell Street, Newport, Rhode Island 02841-1708

Chair's Introduction—2:00

Contributed Papers

2:05

4pEA1. A synthesized tri-modal directional transducer. Alexander L. Butler, John L. Butler, Alan R. D. Curtis (Image Acoust., Inc., 97 Elm St., Cohasset, MA 02025), and Joe Rice (Naval Postgrad. School, Monterey and SSC, San Diego, CA)

A new cylindrical transducer, which achieves directional beams through the synthesis of the omni, dipole, and quadrupole extensional modes of vibration is presented. The three cylindrical radial modes are related in resonant frequency by the formula $f_n = f_0(1+n)^{1/2}$, where $n = 0$, 1, and 2, respectively, and may be excited individually or together by the selection of the circumferential voltage distribution. It will be shown that, in the frequency band between the dipole and quadrupole modes, a simple four-step real voltage distribution can be used to synthesize a smooth transmitting response with constant beamwidth patterns. From this comparatively simple, small diameter transducer, steered beam widths of 90 deg and 10 dB front to back ratios can be obtained. Theoretical and measured transmitting response and beam pattern results on a 4.5-in.-diam. cylindrical transducer design will be presented and compared in the operating frequency range from 15 to 20 kHz. [SBIR work supported by ONR and SPAWAR.]

2:20

4pEA2. Underwater acoustic intensity measurements using a pressure/particle-acceleration intensity vector sensor. Kang Kim, Gerald C. Lauchle, and Thomas B. Gabrielson (Grad. Prog. in Acoust. and Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804)

The acoustic field of simple sources is investigated by measuring the simultaneous, colocated acoustic pressure and particle acceleration with a neutrally buoyant underwater acoustic intensity vector sensor. This sensor consists of a piezoceramic hollow-cylinder pressure transducer, and a pair of miniature accelerometers mounted inside of the cylinder. A free-field calibration is performed under the ice of a flooded quarry, while in a laboratory test tank, the 2-D acoustic intensity field is mapped for various acoustic source configurations. These include a single spherical source and a pair of spherical sources that vibrate in or out of phase with each other. A sinusoidal signal of various frequencies between 2 and 20 kHz is gated in all of these experiments in order to suppress the effects of environmental reflections. The receiving beam pattern of the intensity sensor is also measured and reported. An analytical model for the interaction of two closely spaced spherical radiators is developed, and predictions from it compare well with the measured intensity field. Some limitations in the experiments, and future work are discussed. [Work supported by ONR, Code 321SS.]

2:35

4pEA3. The design, fabrication, and acoustic calibration of a 1-3 piezoelectric polymer composite-based cylindrical array module for towed applications. Kim C. Benjamin and Jeffrey A. Szelag (NAVSEA Div. Newport, 1171 Howell St., Newport, RI 02871)

A cylindrical transducer array module containing eight double-layered 1–3 piezocomposite elements was acoustically calibrated in the frequency range 10–100 kHz. Each wedge-shaped element, covering approximately one octant (45) of the cylindrical radiating aperture, was housed in a molded syntactic foam hub component with its outer surface conforming to a cylindrical aperture 12.7 mm in diameter. The construction approach described would allow one to connect several modules of similar (or varying) diameter along the axial direction for towed-array applications requiring directional transmit and receive performance. The double-layer element design provides two independent electrical control points for optimizing the spatial and spectral response in both transmit and receive modes. The cylindrical module fabrication approach, associated tooling, and measured acoustic results for a single module under various drive conditions will be presented to illustrate the utility of this new transducer technology. [Work sponsored by the U.S. Navy.]

2:50

4pEA4. Measurement of stress/charge coefficients of a thin polyvinylidene fluoride film. Marcellin Zahui and George Bibel (Mech. Eng., Univ. of North Dakota, Grand Forks, ND 58202)

A new experimental/analytical approach is used to measure a piezoelectric coefficient of polyvinylidene fluoride (PVDF). This method is based on the volume displacement of a vibrating beam. The obtained value of the stress/charge coefficient is checked using modal coordinate measurement theory and finite element analysis. A PVDF film is cut into two sensors shape. These sensors measure, respectively, the volume displacement and modal coordinate of a simply supported beam vibrating at its first mode. An accelerometer and a force gauge are used to measure the volume displacement and the modal coordinate of the beam. Then, PVDF volume displacement sensor equation developed by Guigou *et al.* is used with the experimental data to find the stress/charge coefficient e31. This value of e31 is first checked using the PVDF modal sensor equation from Lee and Moon. Then, the ANYS finite element code is used to validate the method. Experimental value of e31 obtained at room temperature is presented and compared with existing data. Excellent agreement is found. The extension of the method to 2D structure to determine the value of e31 and e32 simultaneously is introduced.

3:05

4pEA5. Development of a high sensitive MEMS hydrophone using PVDF. Vijay K. Varadan, Bei Zhu, and Jose K.A (212 EES Bldg., Penn State Univ., University Park, PA 16802)

The design and experimental evaluation of a PVDF-based MEMS hydrophone is presented in this paper. The basic structure of the hydrophone was fabricated on a silicon wafer using standard NMOS process technology. A MOSFET with extended gate electrode was designed as the interface circuit to the sensing material, which is a piezoelectric polymer, polyvinylidene difluoride (PVDF). Acoustic impedance possessed by this piezoelectric material provides a reasonable match to water, which makes it very attractive for underwater applications. The electrical signal generated by the PVDF film was directly coupled to the gate of the MOSFET. To minimize the parasitic capacitance underneath the PVDF film and hence improve the device sensitivity, a thick photoresist was first employed as the dielectric layer under the extended gate electrode. For underwater operation, a waterproof Rho-C rubber encapsulated the hydrophone. A silicon nitride layer passivated the active device, which is a good barrier material to most mobile ions and solvents. The device after passivation also shows a lower noise level. The theoretical model developed to predict the sensitivity of the hydrophone shows a reasonable agreement between the theory and the experiment.

3:20–3:30 Break

3:30

4pEA6. Light-weight low-frequency loudspeaker. Robert Corsaro and James Tressler (Naval Res. Lab., Washington, DC 20375, Corsaro@NRL.Navy.mil)

In an aerospace application, we require a very low-mass sound generator with good performance at low audio frequencies (i.e., 30–400 Hz). A number of device configurations have been explored using various actuation technologies. Two particularly interesting devices have been developed, both using ''Thunder'' transducers (Face Intl. Corp.) as the actuation component. One of these devices has the advantage of high sound output but a complex phase spectrum, while the other has somewhat lower output but a highly uniform phase. The former is particularly novel in that the actuator is coupled to a flat, compliant diaphragm supported on the edges by an inflatable tube. This results in a radiating surface with very high modal complexity. Sound pressure levels measured in the far field (25 cm) using only 200-V peak drive (one-third or its rating) were nominally 74 6 dB over the band from 38 to 330 Hz. The second device essentially operates as a stiff low-mass piston, and is more suitable for our particular application, which is exploring the use of active controlled surface covers for reducing sound levels in payload fairing regions. [Work supported by NRL/ONR Smart Blanket program.]

3:45

4pEA7. Background noise model of the piezoresistive microphone. Allan J. Zuckerwar, Theodore R. Kuhn (NASA Langley Res. Ctr., Hampton, VA 23681), and M. Roman Serbyn (Morgan State Univ., Baltimore, MD 21231, rserbyn@morgan.edu)

A model has been developed to describe the sources of background noise in piezoresistive microphones. Three noise sources have been identified: mechanical Johnson noise due to the Brownian motion of air molecules impinging on the diaphragm, electrical Johnson noise in the bridge resistors, and generation–recombination noise in the bridge resistors. The model includes a set of lumped elements to represent the acoustic impedance of the diaphragm, another set of lumped elements to account for the stress across the bridge resistors, and a controlled source to yield an output voltage proportional to the stress due to the piezoresistive effect. Predictions based on the manufacturer's data agree with experimental measurements. Representing the piezoresistive microphone as a nonreciprocal device, the model can be used to determine the microphone sensitivity and frequency response, as well as the background noise.

4:00

4pEA8. PC-based real-time acoustic source locator and sound capture system for teleconferencing. Ashutosh Morde, Deborah Grove, and Robert Utama (Ctr. for Adv. Information Processing, Rutgers Univ., 96 Frelinghuysen Rd., Piscataway, NJ 08854, amorde@caip.rutgers.edu)

A PC-based real time acoustic source locator and sound capture system has been developed. The system is implemented using Frontier Design A/D converters and the Intel Signal Processing Library directly on a 1 GHz Pentium III machine, without a DSP board. The source locator uses the cross-power spectral phase to locate a moving talker. The algorithm also uses an energy detector that minimizes incorrect location estimates by neglecting frames with high background noise. The source locator provides 8 location estimates per second. A 16-element 0.90 m linear delay-sum beamformer has also been implemented in the system as a method for selective sound capture. The ability of the source locator to detect talkers in a typical office environment is evaluated. In addition, the array response is measured. [Work supported by Intel.]

4:15

4pEA9. An analytical application of a time-reversal mirror in a sensor with solid cylindrical waveguides. Anthony D. Puckett and M. L. Peterson (Dept. of Mech. Eng., Univ. of Maine, 5711 Boardman Hall, Orono, ME 04469, anthony.puckett@umit.maine.edu)

A solid cylindrical waveguide or buffer rod is commonly used to isolate piezoelectric transducers from high temperature and pressure. Due to constraints on the waveguide dimensions, multiple longitudinal modes are often propagated. Dispersion and the existence of the multiple modes make it difficult to perform standard ultrasonic measurements. To simplify the measurements an analytical model was developed to model the dispersion in the multimode waveguide. This model was used to calculate the time-reversed signal required to reconstruct a pulse in the waveguide. The time-reversed signal was used in experiments to demonstrate the ability of the calculated time-reversed signal to construct a pulse. The experiments also showed the ability of the analytical model to represent the dispersion in the waveguide. Results are presented and the limitations of the analytical model presented are discussed. [Research sponsored by the Ballistic Missile Defense Organization through the ONR, Science Officer Dr. Y. D. S. Rajapakse.]

4:30

4pEA10. Implementation of CHIEF in MATLAB for prediction of sound radiated from arbitrarily shaped vibrating bodies. Mark Christensen and Koorosh Naghshineh (Mech. and Aero. Eng. Dept., Western Michigan Univ., Kalamazoo, MI 49008)

The Boundary Element Method can be used to predict sound pressure levels radiated from an arbitrarily shaped vibrating body. Using the direct, indirect, and approximation formulation one can solve such an exterior acoustic problem. In this work, the direct formulation is chosen. The two major methods employed for implementation of this formulation are the Combined Helmholtz Integral Equation Formulation (CHIEF) and the Burton–Miller Method. It was decided that the CHIEF method would be the focus of this research. Using the CHIEF method as a guide, a computer simulation was developed that incorporates quadratic elements, calculates the sound pressure and sound power, is able to import surface velocities, and uses the program IDEAS as a preprocessor. This program was verified using theoretical models as well as experimental measurements conducted in the Western Michigan University Noise and Vibration Laboratory. This program was written in Matlab with the understanding that it can be processed on a personal computer.

THURSDAY AFTERNOON, 6 JUNE 2002 RIVERS ROOM, 1:15 TO 2:30 P.M.

Session 4pMU

Musical Acoustics and Psychological and Physiological Acoustics: Music Recognition Systems II

James W. Beauchamp, Chair
School of Music, Department of ECE, University of Illinois at Urbana-Champaign, 1114 West Nevada, Urbana, Illinois 61801

Contributed Papers

1:15

4pMU1. Independent component analysis for onset detection in piano trills. Judith C. Brown (Phys. Dept., Wellesley College, Wellesley, MA 02181 and Media Lab., MIT, Cambridge, MA 02139), Jeremy G. Todd (iZotope, Inc., Shrewsbury, MA 01545), and Paris Smaragdis (Media Lab., MIT, Cambridge, MA 02139)

The detection of onsets in piano music is difficult due to the presence of many notes simultaneously and their long decay times from pedaling. This is even more difficult for trills where the rapid note changes make it difficult to observe a decrease in amplitude for individual notes in either the temporal wave form or the time dependent Fourier components. Occasionally one note of the trill has a much lower amplitude than the other making an unambiguous determination of its onset virtually impossible. We have analyzed a number of trills from CD's of performances by Horowitz, Ashkenazy, and Goode, choosing the same trill and different performances where possible. The Fourier transform was calculated as a function of time, and the magnitude coefficients served as input for a calculation using the method of independent component analysis. In most cases this gave a more definitive determination of the onset times, as can be demonstrated graphically. For comparison identical calculations have been carried out on recordings of midi generated performances on a Yamaha Disclavier piano.

1:30

4pMU2. Two-voice fundamental frequency estimation. Alain de Cheveigné (CNRS, Ircam, 1 Pl. Igor Stravinsky, 75004 Paris, France, cheveign@ircam.fr)

An algorithm is presented that estimates the fundamental frequencies of two concurrent voices or instruments. The algorithm models each voice as a periodic function of time, and jointly estimates both periods by cancellation according to a previously proposed method [de Cheveigné and Kawahara, Speech Commun. **27**, 175–185 (1999)]. The new algorithm improves on the old in several respects; it allows an unrestricted search range, effectively avoids harmonic and subharmonic errors, is more accurate (it uses two-dimensional parabolic interpolation), and is computationally less costly. It remains subject to unavoidable errors when periods are in certain simple ratios and the task is inherently ambiguous. The algorithm is evaluated on a small database including speech, singing voice, and instrumental sounds. It can be extended in several ways; to decide the number of voices, to handle amplitude variations, and to estimate more than two voices (at the expense of increased processing cost and decreased reliability). It makes no use of instrument models, learned or otherwise, although it could usefully be combined with such models. [Work supported by the Cognitique programme of the French Ministry of Research and Technology.]

1:45

4pMU3. Matching untrained singers using dynamic time warping. Mark Kahrs (Dept. of Elec. Eng., Univ. of Pittsburgh, 348 Benedum Hall, Pittsburgh, PA 15261)

Untrained singers lack precise control of pitch and timing, therefore any successful music search algorithm must accommodate inaccuracies in both dimensions. Furthermore, finding key signatures in untrained singers is not easy (or even plausible in some cases). Contours can be formed by up and down intervals and then matched against the database. An experimental system is described that matches an acoustic vocal input against a library of monophonic musical scores and returns a list of matched results. While partially successful, errors result from the detection of pitch on and off times as well as from pitch detection errors. Pitch detection is easiest during vowels, but the transitions from vowel to consonant can cause problems. In spite of these inaccuracies, the prototype algorithm using dynamic time warping (a dynamic programming algorithm) exceeded 95% accuracy for nine different singers for a simple tune library. However, errors computed by a difference metric can result in close misses, even for very dissimilar songs. A better metric is needed to clearly illustrate song differences.

2:00

4pMU4. Detection and representation of rhythm and tempo I: The rhythmogram. Neil P. M. Todd (Dept. of Psych., Univ. of Manchester, Manchester M13 9PL, UK)

In the paper an algorithm is presented for the computation of a primitive representation of rhythmic structure [N. P. Todd, J. New Mus. Res. **23**, 25–70 (1994)]. The algorithm consists of the following stages: (1) cochlear filtering, by means of an ear model based on the gammatone filter-bank [M. Cooke, *Modeling Auditory Processing and Organisation* (Cambridge University Press, Cambridge, MA, 1992)]; (2) modulation filtering, by means of a Gaussian low-pass modulation filter-bank; (3) peak detection and integration of the output of the modulation filter-bank. When the peak outputs of the modulation filters are plotted in a time-constant/time diagram they form a representation, referred to as a *rhythmogram*, which resembles a more conventional tree diagram of rhythmic structure. A single acoustic event gives rise to a single branch in the representation so that integration along a branch yields a single number corresponding to the perceptual prominence or stress of the event. The integration mechanism can be thought of as an acoustic sensory memory process and accounts for a number of important phenomena in rhythm perception, including persistence, partial backward masking, interval produced accents and grouping by temporal proximity and accent. This representation may form the input to higher level learning and recognition mechanisms.

2:15

4pMU5. Detection and representation of rhythm and tempo II: The modulation spectrogram. Neil P. M. Todd (Dept. of Psych., Univ. of Manchester, Manchester M13 9PL, UK)

In the paper a second algorithm is presented for the computation of a primitive representation of rhythmic and temporal structure [N. P. Todd, C. S. Lee, and D. J. O'Boyle, J. New Mus. Res. **28**, 5–28 (1999)]. The algorithm consists of the following stages: (1) cochlear filtering, by means of an ear model based on the gammatone filter-bank, (2) modulation filtering, by means of a constant-*Q* band-pass modulation filter-bank; (3) peak detection and sensorimotor filtering; (4) feedforward control of a simple plant to simulate tapping. The output of the modulation filter-bank produces a representation in the form of a *modulation spectrogram* whereby any periodic input gives rise to a harmonic series. Metrical structure may be inferred from the harmonic structure of the spectrum with the addition of a learning and recognition mechanism. This algorithm may account for a number of additional phenomena including the approximate Weber law property of tempo sensitivity, adaptation to perturbations in the period or phase and the autocorrelation statistics of sensorimotor synchronization. The performance of the algorithm is evaluated by comparison with the tapping behavior of 20 human subjects to the 48 fugue subjects from the Well-Tempered Clavier by J. S. Bach.

THURSDAY AFTERNOON, 6 JUNE 2002 DUQUESNE ROOM, 2:00 TO 5:00 P.M.

Session 4pNS

Noise: Noise Control Devices—Analysis and Design (Lecture/Poster Session)

Richard D. Godfrey, Chair

Owens-Corning, 2790 Columbus Road, Route 16, Granville, Ohio 43023-1200

Chair's Introduction—2:00

Contributed Papers

2:05

4pNS1. Optimally tuned vibration absorbers to control sound transmission. Michael Grissom, Ashok Belegundu, and Gary Koopmann (Penn State Univ., 157 Hammond Bldg., University Park, PA 16802, mdg184@psu.edu)

A design optimization method is proposed for controlling broadband vibration of a structure and it concomitant acoustic radiation using multiple-tuned absorbers. A computationally efficient model of a structure is developed and coupled with a nonlinear optimization search algorithm. The eigenvectors of the original structure are used as repeated basis functions in the analysis of the structural dynamic re-analysis problem. The re-analysis time for acoustic power computations is reduced by calculating and storing modal radiation resistance matrices at discrete frequencies. The matrices are then interpolated within the optimization loop for eigenvalues that fall between stored frequencies. The method is demonstrated by applying multiple-tuned vibration absorbers to an acoustically-excited composite panel. The absorber parameters are optimized with an objective of maximizing the panel's sound power transmission loss. It is shown that in some cases the optimal solution includes vibration absorbers that are tuned very closely in frequency, thus acting effectively as a broadband vibration absorber (BBVA). The numerical model and design optimization method are validated experimentally, and the BBVA is found to be an effective noise abatement tool.

2:20

4pNS2. Diffraction by jagged edge noise barriers. Jeong Ho You and Penelope Menounou (Univ. of Texas, Austin, TX 78703)

Experimental data have shown that the noise barrier's performance can be improved by introducing jaggedness into the top edge of the barrier. In the present work a combined experimental/numerical investigation was performed. The numerical investigation employs the Directive Line Source Model [Menounou *et al.,* J. Acoust. Soc. Am. **107**, 103–111 (2000)] and reveals the basic global characteristics of a jagged profile that increases the shielding effect over a straight edge barrier having the same average height. For a jagged edge profile consisting of finite-length segments inclined with respect to a straight top edge it was found that its performance depends on the inclination angle of the finite-length segments with respect to the horizontal, the directivity of the diffracted field, and the length of the inclined segments with respect to the spatial duration of the incident sound signal. It was also found that jagged profiles offering an increased shielding effect for certain receiver locations are different from jagged profiles offering moderate improvement over many receiver locations. The numerical predictions were verified experimentally. Experiments were carried out in air using a spark source generating N-shaped pulses, a capacitor microphone and a rigid aluminum plate with various jagged top edge profiles. [Work supported by Texas ATP.]

2:35

4pNS3. Use of tile overlays to reduce the radiation efficiency of radiating surfaces. Gary Koopmann, Eric Salesky, and Weicheng Chen (Penn State, 157 Hammond, University Park, PA 16802)

An intriguing possibility for reducing the radiation efficiency of a structure is to overlay its radiating surfaces with a contiguous set of low mass tiles. Experimental results show that significant reductions in sound power are possible depending on number of tiles, their geometry, attachment methods, mass, stiffness, and the amount of shear damping used to connect their edges to one another. This paper presents results of a combined theoretical and experimental study performed on a clamped, flat plate driven at its center and with nearly infinite baffle boundary conditions. Tiles are overlayed on the surface of the plate and lightly attached with a polymer material. A scanning laser vibrometer is used to examine the vibration patterns of the plate and the corresponding tiles for several different tile configurations over a frequency range 20–1000 Hz. Results indicate that sound power reductions occur when the independent response of each tile combines to produce an overall volume velocity that is lower than that of the plate by itself over the frequency range of interest.

2:50

4pNS4. Design and fabrication of a laboratory device for demonstration of Helmholtz resonators. Marissa Melchior, Zoren Gaspar, and Koorosh Naghshineh (Mech and Aero. Eng. Dept., Western Michigan Univ., Kalamazoo, MI 49008)

There is an increasing demand for noise emission reduction in many applications today. These applications include automotive exhausts, fan intakes and exhausts, and air conditioning ducts among many others. One way to attenuate the noise produced by these applications includes the use of a Helmholtz resonator. A Helmholtz resonator, also referred to as a side branch resonator to a duct, is simply a cavity with a known volume connected via a neck to a duct. A simple equation relates frequency and resonator geometry such that the length and cross section of the resonator neck, as well as the volume and shape of the resonator cavity, affect resonator performance. In practice a resonator is made to target a certain frequency of noise, which is to be suppressed. The noise reduction achieved is dependent on the accuracy of construction of the resonator. A laboratory setup was constructed which incorporated many of the most influential variables. This setup will be used as a demonstration device as well as a laboratory apparatus in the Mechanical Engineering curriculum at Western Michigan University.

3:05

4pNS5. In-line quincke tube muffler. William P. Patrick, Rebecca S. Bryant (United Technologies Res. Ctr.), and Larry E. Greenwald (Independent Contractor)

A unique low-pressure-drop muffler is described which has been designed to attenuate low frequency tonal noise in ducts. Flow through the muffler is divided into two noncommunicating paths in the cylindrical configuration which was designed, built, and tested. Half of the flow is ducted through a straight central annulus and the other half is ducted through a partitioned outer annulus which directs the flow in a spiral flow pattern around the inner annulus. Thus the outer flow has a longer path length and the sound within the outer annulus is phase-delayed relative to the inner flow causing destructive interference between the inner and outer waves with resulting strong attenuation at the tuned frequencies. A procedure will be described for designing a muffler (with flow) to produce high attenuation at the fundamental noise tone and all harmonics (up to the first cross mode). Results will be presented which show that the muffler achieved over 20 dB attenuation for the first five harmonics of the incident noise in a flowing duct.

3:20

4pNS6. Jet noise from ultrahigh bypass turbofan engines. Joe W. Posey, Thomas D. Norum, Martha C. Brown (NASA Langley Res. Ctr., Hampton, VA 23681), and Thonse R. S. Bhat (Boeing Commercial Airplanes, Seattle, WA 98124-2207)

Modern commercial jet transport aircraft are powered by turbofan engines. Thrust from a turbofan engine is derived in part from the exhaust of a ducted fan, which may or may not be mixed with the core exhaust before exiting the nacelle. The historical trend has been toward ever higher bypass ratios (BPRs). The BPR is the ratio of air mass passing through the fan to that going through the core. The higher BPR engines can be more efficient and quieter. In general, a higher BPR results in lower average exhaust velocities and less jet noise. In order to address a scarcity of noise data for BPRs greater than 6, an extensive database collection effort was undertaken using the Jet Engine Simulator in NASA Langley's Low Speed Aeroacoustic Wind Tunnel. Forward flight simulations of Mach 0.1, 0.2, and 0.28 were used with BPRs of 5, 8, 11, and 14. Data was taken over the entire operating line of the simulated engines along with parametric deviations to provide a complete set of sensitivity measurements. The results will be used to develop an empirical jet noise prediction capability for ultrahigh bypass engines.

3:35

4pNS7. On windscreen insertion loss. Mariana Buzduga and Richard J. Peppin (Scantek, Inc., 7060 Oakland Mills Rd., Ste. L, Columbia, MD 21046)

Microphone windscreens are used to attenuate wind noise. They add impedance between the source and microphone, not accounted for when the system is checked by acoustical calibrator or when used in the field. The procedure for the characterization of the attenuation has recently been addressed in ANSI S1.17/2000 part 1. But to date, no windscreens have been tested in accordance with that standard. Partially, the reason is that the precision of the procedure has not been determined. As of January 2002 a round robin is addressing just that issue. By June 2002 the results of the uncertainty determination will be available. In the meantime, we have tested some of the samples used and others in a small chamber approaching free-field conditions. We present the results of the tests of insertion loss (not in accordance with S1.17 because we do not have a reverberation room). We will show that, in some cases, the use of a windscreen can easily change a measurement using type 1 instruments to type 2 or worse. Without knowing information about a particular windscreen, the use of a windscreen in still air can drastically change uncertainty of measurement. In moving air conditions we expect even more severe problems.

Poster Papers

All posters will be on display from 1:30 p.m. to 5:00 p.m. Contributors will be at their posters from 3:50 p.m. to 5:15 p.m. To allow for extended viewing time posters will remain on display until 10:00 p.m.

4pNS8. Automation of calculation and design of means of protection of buildings from transport noise. Igor L. Shubin and Irina L. Sorokina (Dept. of Noise Control, Res. Inst. of Bldg. Phys. (NIISF), 21 Lokomotivny proezd, Moscow, 128239 Russia)

There are stated basic principles of design using the created computer program on calculation and projection of means of protection of the buildings from transport noise. A number of problems have been considered which are connected with the use of the existing methods of evaluation of the conditions of transport noise distribution in the urban area and the effectiveness of protection of buildings and constructions from noise (with the use of the automated calculations and design of noise protection).

4pNS9. Noise reduction of a table saw. John Carlson (School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907)

The National Institute for Occupational Safety and Health (NIOSH) is sponsoring a design project to address the noise levels that commonly exist at construction worksites. Through engineering control, the problem of noise emission from a table saw will be addressed. The noise emitting sources will be pinpointed using a sound pressure level meter. With this knowledge, the next step will be to reduce the sound pressure levels at the noise sources. This will be done by using noise reduction techniques such as insulation, and vibration dampening. The goal is to reduce the noise emission to a level between 85 and 90 dB(A).

4pNS10. Noise source identification and control of a contractor grade table saw. Kristin Bleedorn, Matthew McKee, Dale Yarbough, Jr., Chen Yu, Edward L. Zechmann, and J. Adin Mann III (Aero Eng. and Eng. Mech., Iowa State Univ., 2271 Howe Hall, Ames, IA 50011)

Sponsored by the National Institute for Occupational Safety and Health (NIOSH) as part of their initiative to explore noise reduction strategies for construction equipment, a team of engineering students at Iowa State University studied a contractor grade table saw. Based on standards, published work, and preliminary tests, a repeatable noise measurement procedure was developed for the table saw operation. The wood-feed rate and force were measured. With the saw operating in a standard and consistent manner, noise sources on the saw were identified using sound intensity measurement techniques and through the application of noise control strategies to individual sources. At this stage, noise control strategies, such as enclosing the motor, are effective for noise source identification but not practical. The effectiveness of both approaches to identifying the noise sources will be discussed. Based on rank ordering the contribution of each noise source to the overall sound levels, permanent noise control strategies are suggested.

4pNS11. Noise control of a circular saw. Jeff Okrzesik (School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907)

Construction workers are subjected to high levels of noise pollution on a daily basis. Sound levels of about 85 dB(A) with exposures of 8 hours per day will result in permanent hearing loss after many years [Noise and Hearing Loss, NIH Consens Statement Online **8**(1), 1–24 (1990)]. There are numerous sources contributing to the noise problems. Tools are of particular interest considering the amount of time the average worker uses them. A circular saw is a common tool used daily by many construction workers. The levels of noise produced by a circular saw will be tested in order to evaluate the problem. The noise levels as well as the locations of the noise sources will be evaluated. Several ideas regarding the reduction of sound emissions have been researched. There will be three initial focal points of the investigation. The blade properties, including the material, size and tooth design, are all key elements that will be studied closely. The numerous dynamic parts in the motor are also large sources of noise that may be reduced by increasing the efficiency of the motor. Another vital element to reducing noise involved with a circular saw is to study the interaction noise from the piece being worked on.

4pNS12. Aeroacoustics of microphones. Daniel Finfer (Purdue Univ., 1077 Ray W. Herrick Labs., West Lafayette, IN 47907-1077, finfer@ecn.purdue.edu)

Several commercial options exist for prevention of the pseudo-sound often introduced by a microphone in the presence of a flow field. One such device is a so-called ''slit-tube.'' A slit-tube is a tube approximately the diameter of the microphone with which it is used. The length of such a tube can vary anywhere from 25 cm to 50 cm, and it is most generally fit with a stream-lined noise cone. In theory, a perforated slit along the length of the tube encourages net cancellation of turbulent pressure variations, while not interfering with the acoustic field. This poster compares the effectiveness of two similar slit-tube designs. Each tube was placed inside of a wind-tunnel and exposed to white noise. The effects of increasing flow velocities on the transfer-function between the source and receiver signals are analyzed.

THURSDAY AFTERNOON, 6 JUNE 2002 STERLINGS ROOMS 2 AND 3, 1:00 TO 2:15 P.M.

Session 4pPAa

Physical Acoustics: General Acoustics Modeling

Doru Velea, Cochair
Planning Systems, Incorporated, 12030 Sunrise Valley Drive, Reston Plaza I, Suite 400, Reston, Virginia 20191

Roger M. Waxler, Cochair
National Center for Physical Acoustics, University of Mississippi, Coliseum Drive, University, Mississippi 38677

Contributed Papers

1:00

4pPAa1. Integral equation approach for analyzing acoustic interactions with wrinkled flame fronts. Lei Wu and Tim Lieuwen (School of Aerosp. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0150, lei.wu@ae.gatech.edu)

Flame-acoustic wave interactions play an important role in the characteristic unsteadiness of turbulent combustion processes. They arise because acoustic waves incident from other parts of the flame or from external sources impinge upon the flame and are scattered because of the significant change in sound speed and density at the flame front. The characteristics of the scattered waves are complex, due to the fact that they are interacting with a dynamic flame surface that is convoluted over a broad range of length and time scales. Over a wide range of frequencies (e.g., up to about 20 kHz), such turbulent flames may be treated as nonpassive (i.e., the flame acts as a source of acoustic energy) temperature discontinuities. We describe here an integral equation approach, similar to the development of the Ffowcs Williams–Hawking (FWH) equation, that

4p THU. PM

is suitable for analysis of acoustic-flame interactions. This FWH equation is generalized for the case of interest where the surface is penetrable, not passive, and whose area varies in time. The integration surface consists of the time varying flame front that divides two homogeneous regions. The wave motions in these two regions are coupled by appropriate mass, momentum, and energy conservation conditions.

1:15

4pPAa2. New anisotropic perfectly matched layer boundary condition for elastic waves in solids. Yibing Zheng (Physical Acoust. Lab., Dept. of Mech. Eng., Yale Univ., New Haven, CT 06520-8286), Xiaojun Huang (MIT, Cambridge, MA 02142), and Robert E. Apfel (Yale Univ., New Haven, CT 06520-8286)

The perfectly matched layer (PML) is currently a popular absorbing boundary condition for numerical modeling of electromagnetic, acoustic, and elastic wave problems. Existing PML formulation for elastic waves uses a field splitting scheme, which increases the number of variables in the computation. The new anisotropic elastic wave PML modifies the elasticity or compliance tensors to be anisotropic. It utilizes fewer variables for both isotropic and anisotropic media, so that computing memory and time can be reduced significantly. Its formulations for Cartesian and curvilinear coordinate systems will be discussed. Moreover, this anisotropic PML can be directly used in the finite element method, and easily be implemented in the finite difference time domain method as well. [Work supported by NASA (NAG3-2147) and by NSF (CTS-9870015).]

1:30

4pPAa3. A simple analytical model for prediction of aeroacoustic cavity oscillations. Donald Bliss and Linda Franzoni (Dept. of Mech. Eng., Duke Univ., Durham, NC 27708-0300)

High-speed flow over cavities in vehicle surfaces often produces intense tonal pressure fluctuations. Although this problem has been of concern for decades, especially for aircraft, the underlying physical mechanisms have not been well understood. A fairly simple analytical model has been developed that improves understanding of the physical mechanisms of pressure fluctuations in shallow rectangular cavities exposed to high-speed flow. The new model describes the important features of the phenomena, and gives good agreement with experimental data. The analysis is constructed by combining waves that can exist in a system having a finite thickness shear layer dividing two acoustic media, one at rest and one in motion. The wave types for this configuration can be interpreted in physical terms. A new aspect of the analysis shows the important effect of shear layer thickness on the convective waves. Fundamentally, only two convective waves and three acoustic waves suffice to represent the phenomenon. The analysis shows the important role of mass addition and removal at the rear of the cavity due to shear layer oscillation. Certain conditions on energetics and shear layer motion at the rear of the cavity must be satisfied simultaneously for oscillation to occur.

1:45

4pPAa4. On the effect of viscosity in scattering from partially coated infinite cylinders. Jerry H. Ginsberg (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

Partial coating of a submerged object introduces discontinuities in the surface properties that complicate an acoustic scattering analysis. Some analyses, such as the work by Feit and Cuschieri [J. Acoust. Soc. Am. **92**, 1721–1728 (2000)], have used weighted residual techniques based on global series representations, thereby generating a solution that satisfies discontinuous boundary conditions in an average sense. Others, such as Ferri, Ginsberg, and Rogers [J. Acoust. Soc. Am. **92**, 1721–1728 (1992)] have used collocation-type formulations. Another issue arises from the fact that the normal velocity changes discontinuously at an impedance discontinuity. The associated large velocity gradients give rise to concern that viscous shear might play a role. The present work considers the two-dimensional problem of scattering of a plane wave incident on an infinite cylinder that is coated with strips of pressure-release material extending over quadrants on the illuminated and shadowed sides, with the remainder of the surface considered to be rigid. The analysis represents the scattered field in terms of acoustic and vortical contributions, both of which are represented in series form. Solutions obtained by the method of weighted residuals are shown to be less accurate than those obtained from a collocation procedure. The results for surface pressure and farfield directivity indicate that viscous effects are important only if the Reynolds number is extremely small.

2:00

4pPAa5. Acoustic field in a cased well with a sectorial crossing channel. Weijun Lin, Chengyu Zhang, Xiuming Wang, and Hailan Zhang (Inst. of Acoust., Chinese Acad. of Sci., P.O. Box 2712, Beijing 100080, PROC, weijun_l@hotmail.com)

The crossing channel in a cased well is a vertical and sectorial column filled with fluid in the cement layer between the steel pipe and formation sometimes formed during well cementing, which is harmful to the well and must be detected. To study the possibility of detecting the crossing channel by using a traditional logging tool, an acoustic field generated by a monopole source in a cased well with a crossing channel of various angles is simulated by a 2.5-D finite difference method for the first time. The results of the 2.5-D finite-difference method for the cases of 0 or 360 deg are compared and consistent with that of the well-known real axis integration method. Snapshots of normal stress and synthetic time-domain waveforms are displayed. The 2-D spectrum in wave number and frequency domain is also calculated, where the influence of the channel is clearer than that in the waveforms. Numerical study demonstrated that a crossing channel can be detected and sized if its angle is greater than 30 deg, and larger offset and lower frequency source are favorable to detect and size the crossing channel. [Work supported by NSFC 19974059.]

Session 4pPAb

Physical Acoustics and Education in Acoustics: Demonstrations in Physical Acoustics

Matthew E. Poese, Chair
Graduate Program in Acoustics, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16801

The Technical Committee on Physical Acoustics and the Committee on Education in Acoustics are proud to present a showcase of demonstration apparati that promises to inspire and challenge your understanding of acoustics. Among the demonstrations are the following:

Acoustic Log Starter: Like a fire starter for a fireplace, a steel tube with a line of small holes drilled along the top is filled with propane and the gas jets are set afire. When the tube is insonified by a loudspeaker placed at one end, the flames above the holes are seen to vary in height along the tube due to second order acoustic effects in the gas contained in the tube.

Longitudinal Waves on a String: A ''Long Stringed Instrument'' will be played. The instrument generates interesting sounds when the steel piano wire strings are excited with longitudinal waves instead of the usual transverse waves.

High Amplitude Waves on a String: Nonlinear waves on a string show ''mode jumping effects.'' The string is a very stretchy light elastic cord and the amplitude can be made quite large.

Standing Waves on a Drum Head: A thin latex sheet is stretched on a circular frame and excited into motion by a loudspeaker placed below the membrane. Beautiful standing wave shapes are stroboscopically illuminated when the drive frequency of the loudspeaker is coincident with a mode of the membrane.

Listening to Bubbles: Ambient noise in the ocean is due in part to bubbles. Small, individual bubbles are entrained when rain drops hit the ocean surface and clouds of bubbles are created under breaking waves. The unique sounds of these events will be demonstrated.

Singing Beer Glass: Bubbles can greatly affect the frequency dependent speed and attenuation of sound. This will be demonstrated using a tube full of bubbles and a glass of beer.

Cavitation Stick: Collapsing vapor bubbles are responsible for damage to pumps and propellers in a process known as cavitation. The phenomenon will be demonstrated in dramatic fashion using a simple device called a ''cavitation stick.''

Helmholtz Resonator Jets: Not only can a Helmholtz resonator provide a jet with enough force to blow out a match, but it can also provide the motive force for an acoustic motor. Four Helmholtz resonators arrayed on a crossbar and balanced at the center of the cross are made to spin around the balance point by creating a high-intensity, two-dimensional standing wave field near the resonators.

Spinning Cups with Sound: After creating a high amplitude two-dimensional standing wave field, first one then four upside down styrofoam coffee cups can be made to spin individually about a balance point. By changing the relative phases of the field, the cups can be made to change their rotational direction.

Acoustic Levitation: Thin sheets of foam are levitated in an intense standing wave field. This effect can be explained with the Bernoulli principle due to large particle velocities at the pressure nodes of the standing wave. Oscillation of the trapped material about its levitation position exhibits an interesting instability.

Acoustic Fountain: A horizontal tube half-filled with water displays miniature fountains when excited with a standing wave. The fountains appear at pressure nodes due to the low-pressure cells caused by the radiation pressure at the pressure anti-nodes.

Singing Flames: An empty tube can be made to resonate at its fundamental frequency when suspended over a flame from a Bunsen burner.

Rijke Tube: A tube having a short stack of screens across its diameter, located between a pressure anti-node and node, is made to resonate when the screens are heated. These oscillations require a steady air current through the tube (provided in this case by natural convection) and continue until the screens cool.

Acoustic Land Mine Detection: A demonstration of the process of imaging buried objects will be shown.

The collaborators for this show, in alphabetical order are:

Anthony A. Atchley, Kevin J. Bastyr, Steven L. Garrett, Robert M. Keolian, Murray S. Korman, Julian D. Maynard, John MacGillivray, Matthew E. Poese, James M. Sabatier, Charles R. Thomas, Ray Scott Wakeland, Preston S. Wilson

Session 4pSA

Structural Acoustics and Vibration and Noise: Active Control of Ducted Fan Noise and Other Active Control Issues

Scott D. Sommerfeldt, Chair
Department of Physics and Astronomy, Brigham Young University, P.O. Box 24673, Provo, Utah 84602-4673

Chair's Introduction—2:00

Invited Papers

2:05

4pSA1. Error signal sensing for active noise control in turbofan engine nacelles. Bruce Walker, Alan Hersh (Hersh Walker Acoust., 780 Lakefield Rd., Unit G, Westlake Village, CA 91361), and Edward Rice (E. J. Rice Consulting, 3116 Dover Ctr. Rd., Westlake, OH 44145)

Blade passage harmonic tones are a significant component of turbofan engine noise under approach conditions. Active noise control (ANC) offers a tool for suppression of these tones, with the caveat that interference patterns in the sound field can lead to poor or even negative correlation between in-duct sound pressures and radiated sound power. Various modal decomposition and wave-number separation schemes have been investigated in attempt to overcome this obstacle. The paper will discuss methods that have been applied to design and implementation of microphone arrays and signal processing techniques to provide unambiguous error signals for ANC systems in these complex acoustical environments and will present results of computer simulations and field tests. [Work sponsored by NASA Glenn and NASA Langley Research Centers.]

2:30

4pSA2. Numerical model for multi-mode active control of turbofan noise. Laralee G. Ireland and Scott D. Sommerfeldt (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602, ireland@byu.edu)

A numerical model has been developed to investigate the possibility of implementing active control to minimize noise radiation from turbofan engines. Previous experimental work on the NASA Glenn Research Center ANC fan have been encouraging, but the question remains of whether the modal approaches investigated can be expected to work effectively on real engines. The engine model presented here uses an indirect boundary element method, implemented with Sysnoise, and a multi-mode Newton algorithm to simulate the active control. Experimental results obtained using the NASA Glenn ANC fan, which has a simplified cylindrical engine geometry, were used as a benchmark for the model. The model results have been found to compare well with the experimental results. Recently, a more realistic, but still axisymmetric, engine geometry has been used. Preliminary results will be presented which indicate that a modal ANC approach should still be effective for controlling turbofan noise.

2:55

4pSA3. Fan noise control using Herschel–Quincke resonators on a production turbofan engine. Ricardo A. Burdisso (Vib. and Acoust. Labs., Virginia Tech, 153 Durham Hall, Blacksburg, VA 24061) and Carl H. Gerhold (NASA Langley Res. Ctr., Hampton, VA 23681-2199)

The Herschel–Quincke (HQ) resonator concept is an innovative technique that consists of installing circumferential arrays of HQ waveguides around the inlet of a turbofan engine. An HQ waveguide is essentially a hollow side tube that travels along (but not necessarily parallel to) the engine axis and attaches to the inlet at each of the two ends of the tube. To investigate the potential of the concept, the approach was tested on a full-scale production Honeywell TFE731-60 engine. An HQ–inlet system containing two arrays was designed to attenuate the blade passage frequency (BPF) tone at approach condition, i.e., 60% engine power. However, the system was tested over the full range of engine power settings. The effects of each array both individually and together were evaluated as compared to the hard-wall case. Both far-field and induct data were recorded during the tests. The results show good attenuation of both the BPF tone and broadband components. Furthermore, reduction of ''buzz-saw'' tones, i.e., additional tones radiated from the inlet when the fan-tip speed goes supersonic, was observed with the HQ system. Some fan distortion effects and increase in noise was observed at higher engine speeds. [Work supported by NASA Langley Research Center.]

3:20–3:30 Break

3:30

4pSA4. Development of shaped PVDF sensors for the measurement of volume displacement. Brian Zellers, Randall Rozema, and Koorosh Naghshineh (Mech. and Aeronautical Eng. Dept., Western Michigan Univ., Kalamazoo, MI 49008)

To attenuate the sound radiated from a vibrating structure, active control can be implemented. When this radiated sound is in the lower frequencies, a reduction of the surface volume displacement of the vibrating structure is suitable for achieving attenuation within these lower frequencies. The sound power emitted from the surface is directly related to the surface volume displacement of the vibrating structure. Traditionally, measurement of the vibration of the structural surface is done using multiple point sensors. Previous work has shown that for a one-dimensional vibrating structure a single sensor made of polyvinylidene fluoride (PVDF) can be used in place of point sensors. Additionally, PVDF sensors have been developed to sense the surface volume displacement in a local area of interest by spanning only that area. However, it has been observed that at higher frequencies both the total and local sensors tend to exhibit reduced accuracy. Thus, a series of experiments was conducted to determine if the PVDF stress/charge coefficient varies as a function of frequency. Additionally, a simulation program was developed to predict sensor charge output for arbitrarily shaped sensors. This simulation, which allows us to develop sensor shapes and predict sensor placement errors, was verified experimentally.

3:45

4pSA5. Decentralized structural acoustic control of a launch vehicle payload fairing. Kenneth D. Frampton (Dept. of Mech. Eng., Vanderbilt Univ., VU Station B 351592, Nashville, TN 37235)

The development of smart structures and active noise and vibration control technologies promised to revolutionize the design, construction and, most importantly, the performance of many complex engineering. However, the early promise of these technologies has not been realized in large-scale systems primarily because of the excessive complexity, cost, and weight associated with centralized control systems. Now, recent developments in MEMS sensors and actuators, along with networked embedded processor technology, have opened new research avenues in decentralized controls. Such a control system consists of numerous nodes, possessing limited computational capability, sensors, and actuators. Each of these nodes is also capable of communicating with other nodes via a wired or wireless network. This results in a dramatic shift in the control system paradigm from that of a single, centralized computer to that of numerous decentralized, networked processors. This work describes the application of such a control system to the reduction of structural acoustic radiation in a launch vehicle payload fairing. A JAVA-based simulation tool is employed to simulate the interactions of the physical system with the networked embedded controllers. Results will indicate the potential for such a control system as well as the limitations imposed by the networked embedded processor hardware.

4:00

4pSA6. A comparison of variable structure control strategy to classical control strategies for active vibration control of distributed structures. Jiaqiang Pan (Dept. of Instrumental Sci. & Eng., Zhejiang Univ., Hangzhou 310027, PROC) and Shuwen Pan (Univ. of California, Irvine, CA 92697)

Independent Modal Space Control (IMSC) is known as an essential method of active vibration control of distributed elastical structures. However, Meirovitch, who advanced the method at the beginning of the 1980s, and other researchers have only investigated modal control law and control performance due to employing classical linear control strateries (e.g., pole allocation, optimal control, etc.) since the 1980s. In this study, authors theoretically and experimentally investigate the modal control law and control performance as using a nonlinear control strategy called variable structure control (VSC) in IMSC of the distributed structures, and then give a comparison of them to the results from using classical control strategies. Besides strong stablity robustness of VSC, the theoretical analysis also shows that, when the sensor–actuator couples employed are less than the number of the modes one desires to control, use of VSC can efficiently avoid the control spillover that often happens at residual modes. For active vibration control of the elastical structures with the classical control strategies, the control spillover usually is a difficult problem to deal with. [Work supported by Zhejiang STPF, PROC.]

4:15

4pSA7. Multi-channel active control of axial cooling fan noise. Kent L. Gee and Scott D. Sommerfeldt (Dept. of Phys. and Astron., N-283 ESC, Brigham Young Univ., Provo, UT 84602, klg33@email.byu.edu)

A multi-channel active control system has been applied to the reduction of free-field tonal noise from a small axial cooling fan. The experimental apparatus consists of an aluminum enclosure which houses the fan, an infrared detector-emitter pair which serves as a reference sensor, loudspeakers, microphones, and appropriate filters and amplifiers. The research fan and loudspeakers have been modeled theoretically as point sources to derive a mathematical expression for radiated power. The minimization of this power has served to guide the number and location of control sources, as well as to discover potential error microphone locations in the extreme near-field. Experiments with these microphone positions have shown that there are predictable near-field locations which consistently lead to significant reductions in the global mean-squared pressure for the first four harmonics of the blade passage frequency. For example, a four channel configuration with the error microphones placed appropriately results in global mean-squared pressure reductions on the order of 10 dB or more for the targeted frequencies.

4:30

4pSA8. Design of model reduction and robust controller for underwater echo cancellation. Woosuk Chang, Vijay K. Varadan, and Vasundara V. Varadan (212 EES Bldg., Penn State Univ., University Park, PA 16802)

This paper presents the design and development of a simple and robust control system for structural acoustics based on state space control theory. The controller was developed using model Hankel singular value (MHSV)-based model reduction, and analog controllers are fabricated. MHSV is developed for a model coordinate system describing vibrating structures, and analog controllers were designed from transfer functions of controllers, which have great advantage over digital controllers in terms of simplicity and cost. This controller design for structural acoustics is applied for controlling sound transmission and reflection. MHSV shows good model reduction of finite-element-based structure models. The feasibility of this controller for the underwater active reflection cancellation is also studied. 1-3 piezocomposite material is used as a transducer that generates the acoustic signal. Its electroacoustic model is developed using the Mason circuit and the control parameters are directly derived. The controller has been designed accordingly and its performance is evaluated using pulse tube. The simulation and measurement show the possibility of achieving a very good reflection reduction in a desired frequency band.

Session 4pSC

Speech Communication and Psychological and Physiological Acoustics: Future of Infant Speech Perception Research: Session in Memory of Peter Jusczyk

Rochelle S. Newman, Cochair
Department of Hearing and Speech Science, University of Maryland, College Park, Maryland 20742

Ruth J. Tincoff, Cochair
Psychology Department, Harvard University, 970 William James Hall, 33 Kirkland Street, Cambridge, Massachusetts 02138

Chair's Introduction—1:00

Invited Papers

1:05

4pSC1. Word recognition and phonetic structure acquisition: Possible relations. James Morgan (Cognit. & Linguistic Sci., Brown Univ., Box 1978, Providence, RI 02912)

Several accounts of possible relations between the emergence of the mental lexicon and acquisition of native language phonological structure have been propounded. In one view, acquisition of word meanings guides infants' attention toward those contrasts that are linguistically significant in their language. In the opposing view, native language phonological categories may be acquired from statistical patterns of input speech, prior to and independent of learning at the lexical level. Here, a more interactive account will be presented, in which phonological structure is modeled as emerging consequentially from the self-organization of perceptual space underlying word recognition. A key prediction of this model is that early native language phonological categories will be highly context specific. Data bearing on this prediction will be presented which provide clues to the nature of infants' statistical analysis of input.

1:25

4pSC2. The future of infant speech perception research: Gotta wear shades. Amanda C. Walley (Dept. of Psychol., Univ. of Alabama, Birmingham, Birmingham, AL 35294)

The future of infant speech perception research is very bright in large part because of the legacy left to us by Peter W. Jusczyk. I will selectively review a number of studies, many of them conducted by Peter and his colleagues, that I consider to be especially interesting and/or provocative and that have helped to advance our understanding of the development of speech perception. For example, in two fairly recent studies, 4.5 to 6-month-old infants were shown to display some incipient spoken word recognition ability [D. R. Mandel *et al.*, Psychol. Sci. **6**, 314–317 (1995); R. Tincoff and P. W. Jusczyk, *ibid.* **10**, 172–175 (1999)]. In effect, these studies have smashed the sound-meaning barrier that was, according to traditional estimates, placed somewhere between 9 and 13 months of age and have important implications for current developmental theory. This work is also illustrative of one trend apparent in recent research, namely, to accord greater attention to infants' perception of ecologically-relevant stimuli and/or their perception of speech stimuli in ecologically-relevant contexts. Still, there is much to be learned about spoken word recognition beyond infancy and before adulthood if we are to arrive at a complete picture of this core aspect of spoken language processing.

1:45

4pSC3. From speech perception to word learning and beyond. Janet F. Werker (Dept. of Psych., 2136 West Mall, Univ. of British Columbia, Vancouver, BC V6T 1Z4, Canada)

From his undergraduate days at Brown, Peter Jusczyk revolutionized our understanding of the link between infant speech perception and language use. This talk reviews how Jusczyk's work influenced my research in three important domains. The talk begins with a discussion of Jusczyk's early work on infant speech perception, and illustrates how that led to my initial work on infant cross-language perception. This is followed by a discussion of Jusczyk's work on sensitivity to probabilistic information and how this influenced the more recent work [Maye, Werker, and Gerken] on the mechanisms underlying changes in phonetic perception. The third research section briefly describes how Jusczyk's work on word segmentation influenced ongoing research in the laboratory on early word learning. In the final few minutes of the talk, an attempt will be made to give a sense of the enormous influence Peter Jusczyk's work had, and continues to have, on our field.

2:05

4pSC4. The development of perceptual attention and articulatory skill in one or two languages. Carol Fowler and Catherine Best (Haskins Labs., 270 Crown St., New Haven, CT 06511)

Infants acquire properties of their native language especially during the second half of the first year of life. Models, such as Jusczyk's WRAPSA, Best's PAM, Kuhl's NLM, and Werker's account describe changes in perceptual or attentional space that may underlie the perceptual changes that infants exhibit. Unknown is the relation of these changes to changes in speechlike vocalizations that occur at the same time. Future research should address whether the perceptual models predict production learning. Other issues concern how the perceptual and articulatory systems develop for infants exposed to more than one language. Do multiple perceptual spaces develop, or does one space accommodate both languages? For infants exposed to just one language, but living in an environment where the ambient and pedagogical language is different (say, infants in a monolingual Spanish home in the U.S.), early language learning fosters learning the native language, but it may impede learning the ambient language. How much or how little does early exposure to the ambient language allow development of perceptual and articulatory systems for the ambient language? A final issue addresses whether the emergence of lexical, morphological and/or syntactic abilities in the second year is related to further changes in speech perception and production. [Work supported by NICHD.]

2:25–2:50
Discussion
Sandra Trehub, Discussion leader

2:50–3:00 Break

Contributed Papers

3:00

4pSC5. Talker variation and word learning. George Hollich III, Peter Jusczyk (Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218), and Michael Brent (Washington Univ.)

While infants must go beyond talker-specific information in recognizing a given word, regardless of the talker, they must also process talker-specific information in order to extract meaning from a particular sound source. Otherwise, for example, they could never recognize whether [hct] referred to a talker's pronunciation of hot, hut, or even hat. This poster suggests that not only do infants process talker-specific information, but they also make use of it both to extract invariant properties in learning a new word and in recognizing talker-specific tokens faster. Using the splitscreen preferential looking paradigm, two studies were conducted that examined how talker-specific properties and variation among talkers could facilitate word learning. Results of study 1 indicated that word learning was facilitated in the case where infants heard different talkers. Thus, talker variation is critical for the extraction of invariant properties of a word. However, the results of study 2 indicated that talker-specific properties were encoded and used to help these infants recognize and learn the referents of these words. Given this evidence, it is suggested that infants appear to be using talker-specific information to form abstract representations of the invariant properties of words.

3:15

4pSC6. The role of language experience in word segmentation: A comparison of English, French, and bilingual infants. Linda Polka, Megha Sundara, and Stephanie Blue (School of Commun. Sci. and Disord., McGill Univ., Montreal, QC H3G 1A8, Canada)

English-learning and French-learning 7.5-month-old infants were tested using the headturn preference procedure on their ability to segment bisyllabic words in both English and French. In the English condition infants were familiarized with trochaic bisyllables, the predominant stress pattern found in English, and then presented English passages with and without the familiarized words. In the French condition, infants were familiarized with iambic bisyllables, the characteristic word stress pattern found in French, and were then presented French passages with and without the familiarized words. Findings indicate that by 7.5 months of age, infants' learning either a syllable-timed (French) or a stressed-timed language (English) can segment bisyllabic words with the predominant stress pattern of their native language. However, French infants fail to segment English trochaic words from English passages; data on English infants' segmentation in the French condition are forthcoming. If both groups fail to segment in a rhythmically different non-native language, it will confirm that word segmentation abilities of 7.5-month-old infants are influenced by the prosodic structure of the native language. Preliminary results obtained from infants who are regularly exposed to both languages will also be reported.

3:30

4pSC7. Language discrimination without language: Experiments on tamarin monkeys. Ruth Tincoff, Marc Hauser, Geertrui Spaepen (Dept. of Psych., Harvard Univ., 33 Kirkland St., Cambridge, MA 02138, tincoff@wjh.harvard.edu), Fritz Tsao (Hillsdale College, Hillsdale, MI 49242), and Jacques Mehler (LSCP, 75006 Paris, France)

Human newborns can discriminate spoken languages differing on prosodic characteristics such as the timing of rhythmic units [T. Nazzi *et al.*, JEP:HPP **24**, 756–766 (1998)]. Cotton-top tamarins have also demonstrated a similar ability to discriminate a morae- (Japanese) vs a stress-timed (Dutch) language [F. Ramus *et al.*, Science **288**, 349–351 (2000)]. The finding that tamarins succeed in this task when either natural or synthesized utterances are played in a forward direction, but fail on backward utterances which disrupt the rhythmic cues, suggests that sensitivity to language rhythm may rely on general processes of the primate auditory system. However, the rhythm hypothesis also predicts that tamarins would fail to discriminate languages from the same rhythm class, such as English and Dutch. To assess the robustness of this ability, tamarins were tested on a different-rhythm-class distinction, Polish vs Japanese, and a new same-rhythm-class distinction, English vs Dutch. The stimuli were natural forward utterances produced by multiple speakers. As predicted by the rhythm hypothesis, tamarins discriminated between Polish and Japanese, but not English and Dutch. These findings strengthen the claim that discriminating the rhythmic cues of language does not require mechanisms specialized for human speech. [Work supported by NSF.]

3:45

4pSC8. Infants long-term memory for complex music. Beatriz Ilari (Faculty of Music, McGill Univ., 555 Sherbrooke St. W., Montreal, QC H3A 1E3, Canada, bilari@po-box.mcgill.ca), Linda Polka (School of Commun. Sci. and Disord., McGill Univ.), and Eugenia Costa-Giomi (Faculty of Music, McGill Univ.)

In this study we examined infants' long-term memory for two complex pieces of music. A group of thirty 7.5 month-old infants was exposed daily to one short piano piece (i.e., either the Prelude or the Forlane by Maurice Ravel) for ten consecutive days. Following the 10-day exposure period there was a two-week retention period in which no exposure to the piece occurred. After the retention period, infants were tested on the Headturn Preference Procedure. At test, 8 different excerpts of the familiar piece were mixed with 8 different foil excerpts of the unfamiliar one. Infants showed a significant preference for the familiar piece of music. A control group of fifteen nonexposed infants was also tested and showed no preferences for either piece of music. These results suggest that infants in the exposure group retained the familiar music in their long-term memory. This was demonstrated by their ability to discriminate between the different excerpts of both the familiar and the unfamiliar pieces of music, and by their preference for the familiar piece. Confirming previous findings (Jusczyk and Hohne, 1993; Saffran *et al.*, 2000), in this study we suggest that infants can retain complex pieces of music in their long-term memory for two weeks.

4:00

4pSC9. Application of a model of the auditory primal sketch to cross-linguistic differences in speech rhythm: Implications for the acquisition and recognition of speech. Neil P. M. Todd (Dept. of Psych., Univ. of Manchester, Manchester M13 9PL, UK) and Christopher S. Lee (Univ. of Manchester, Manchester M13 9PL, UK)

It has long been noted that the world's languages vary considerably in their rhythmic organization. Different languages seem to privilege different phonological units as their basic rhythmic unit, and there is now a large body of evidence that such differences have important consequences for crucial aspects of language acquisition and processing. The most fundamental finding is that the rhythmic structure of a language strongly influences the process of spoken-word recognition. This finding, together with evidence that infants are sensitive from birth to rhythmic differences between languages, and exploit rhythmic cues to segmentation at an earlier developmental stage than other cues prompted the claim that rhythm is the key which allows infants to begin building a lexicon and then go on to acquire syntax. It is therefore of interest to determine how differences in rhythmic organization arise at the acoustic/auditory level. In this paper, it is shown how an auditory model of the primitive representation of sound provides just such an account of rhythmic differences. Its performance is evaluated on a data set of French and English sentences and compared with the results yielded by the phonetic accounts of Frank Ramus and his colleagues and Esther Grabe and her colleagues.

4:15–5:00
Panel Discussion

THURSDAY AFTERNOON, 6 JUNE 2002 BENEDUM ROOM, 2:00 TO 5:10 P.M.

Session 4pSP

Signal Processing in Acoustics and Underwater Acoustics: Advances in Sonar Imaging Techniques Including Synthetic Aperture Sonar and Computed Tomography

Brian G. Ferguson, Cochair
Maritime Systems Division, Defence Science and Technology Organization, P.O. Box 44, Pyrmont 2009, Australia

Kerry W. Commander, Cochair
Coastal Systems Station, Code R21, Panama City, Florida 32407-7001

Chair's Introduction—2:00

Invited Papers

2:05

4pSP1. Imaging sonar development for mine countermeasure applications. Bruce Johnson (Office of Naval Res., 800 N. Quincy St., Arlington, VA 22217)

Over the past decade the Office of Naval Research (ONR) has sponsored research to improve the mine countermeasures (MCM) community's ability to detect, classify, reacquire, and identify mines. In the first stage of wide area detection and classification, the single biggest problem is the ability to distinguish mines from clutter with side-scan sonar assets. In the reacquire-identify stage, divers employ a single-beam aural sonar for guidance to close proximity for visual identification of targets. Current sonar research efforts in wide-area search focus on clutter rejection through high-resolution imaging by synthetic aperture processing and cuing from broadband response characteristics. For reacquisition and identification, research has focused on compact, multibeam imagers. ONR's focus on transition from towbodies and divers to numbers of small, unmanned underwater vehicles (UUVs) to improve MCM clearance rates has had a major impact on sensor development as well. The presentation will provide an introduction to the session by providing a brief history of the development of navy imaging sonars, how they are employed, and an overview of the current thrusts that will lead to the next generation of Navy MCM sonars.

2:30

4pSP2. Application of computed tomography to the acoustic imaging of sea mines at safe standoff distances. Ron Wyber (Midspar Systems Pty. Ltd., 24 Farrer Pl., Oyster Bay NSW 2225, Australia) and Brian Ferguson (Defence Sci. and Technol. Organisation, P.O. Box 44, Pyrmont NSW 2009, Australia)

Typical imaging sonars make use of very high frequencies and arrays with a large number of sensors to provide high angular resolution. With such arrays the spatial resolution degrades and high propagation losses occur as the sonar is moved away from the target being imaged and operational ranges are generally 5 m or less. An alternative imaging technique has been implemented using computed tomography in which the target is insonified with a broad beam from multiple aspects. The spatial resolution achievable with this method is determined only by the bandwidth of the test signal and not by the beamwidth of the sonar or the distance of the sonar from the target. Results are presented showing clear images of mines and mine-like objects in which a spatial resolution of better than 2 cm is achieved at a range of 60 m, even for low target strength mines. Features are evident in the images that are related to both the internal and external shape of the targets as well as structural waves propagating within the targets. This offers the potential to provide unambiguous classification of mines from long ranges, even in a harsh operational environment.

2:55

4pSP3. Recent advances in synthetic aperture sonar for proud and buried object detection. M. Pinto, A. Bellettini, R. Hollett, and L. Wang (NATO SACLANT Undersea Res. Ctr., La Spezia, Italy)

This paper presents a comprehensive research program into synthetic aperture sonar (SAS) for proud and buried object detection. Experimental results are presented, which demonstrate the accuracy achievable with data-driven micronavigation (motion estimation) techniques, such as the displaced phase center antenna (DPCA) as well as aided inertial navigation systems (AINS). The combination of both methods is shown to provide a marked improvement in SAS performance compared with either method used alone. The impact of shallow and very shallow water environments is also discussed, using experiments specifically designed to study the effect of multipath and acoustic fluctuations. Finally, the issue of fast SAS image formation algorithms is addressed in the context of wideband/widebeam systems, nonlinear platform tracks, and sound velocity changes (as when sound penetrates into sediments). Preliminary results are presented for fast-factorized backprojection (FFBP), a promising algorithm derived for low-frequency SAR applications. [This work is done within the framework of a multinational Joint Research Program on Mine Detection and Classification and the authors would like to acknowledge the participation of all the Nations involved, in particular the UK and Norway.]

3:20

4pSP4. Synthetic aperture sonar imaging in a random ocean. Enson Chang (Dynam. Technol., Inc., 21311 Hawthorne Blvd., Ste. 300, Torrance, CA 90503-5602)

The maximum achievable aperture of a coherent imaging sonar is normally considered to be limited by the horizontal coherence length of the medium, which is on the order of several tens to several hundred wavelengths. Synthetic aperture sonar results from the past several years, however, have clearly exceeded this limit by an order of magnitude or more. It appears that the horizontal coherence length alone does not describe some important aspects of the phase errors present in real data. In particular, we suggest that underlying structures in the phase errors are exploitable in processing algorithms and are responsible for these results. We give a number of numerical examples from sub-kHz to tens of kHz, including new results on subcritical imaging of buried objects.

3:45–4:00 Break

4:00

4pSP5. Motion compensation technique for wide beam synthetic aperture sonar. Jose E. Fernandez, Daniel A. Cook, and James T. Christoff (Naval Surface Warfare Ctr., Coastal Systems Station, Dalhgren Div., 6703 W. Hwy. 98, Panama City, FL 32407)

Optimal performance of synthetic aperture sonar (SAS) systems requires accurate motion and medium compensation. Any uncorrected deviations from those assumed during the SAS beam formation process can degrade the beam pattern of the SA in various ways (broadening and distortion of the main lobe, increased side lobes and grating lobes levels, etc.). These would manifest in the imagery in the form of degraded resolution, blurring, target ghosts, etc. An accurate technique capable of estimating motion and medium fluctuations has been developed. The concept is to adaptively track a small patch on the sea bottom, which is in the order of a resolution cell, by steering the SAS beam as the platform moves in its trajectory. Any path length differences to that patch (other than the quadratic function product of the steering process) will be due to relative displacements caused by motion and/or medium fluctuations and can be detected by cross-correlation methods. This technique has advantages over other data driven motion compensation techniques because it operates in a much higher signal-to-noise beam space domain. The wide beam motion compensation technique was implemented in MATLAB and its performance evaluated via simulations. An overview of the technique, simulation, and results obtained are presented.

4:25

4pSP6. Acoustical imaging of high-frequency elastic responses of targets. Scot F. Morse (Div. of Computer Sci., Western Oregon Univ., Monmouth, OR 97361), Brian T. Hefner (Univ. of Washington, Seattle, WA 98105-6698), and Philip L. Marston (Washington State Univ., Pullman, WA 99164-2814)

Acoustical imaging was used to investigate high-frequency elastic responses to sound of two targets in water. The backscattering of broadband bipolar acoustic pulses by a truncated cylindrical shell was recorded over a wide range of tilt angles [S. F. Morse and P. L. Marston, ''Backscattering of transients by tilted truncated cylindrical shells: time-frequency identification of ray contributions from measurements,'' J. Acoust. Soc. Am. (in press)]. This data set was used to form synthetic aperture images of the target based on the data within different angular apertures. Over a range of viewing angles, the visibility of the cylinder's closest rear corner was significantly enhanced by the meridional flexural wave contribution to the backscattering. In another experiment, the time evolution of acoustic holographic images was used to explore the response of tilted elastic circular disks to tone bursts having frequencies of 250 and 300 kHz. For different tilt angles, specific responses that enhance the backscattering were identified from the time evolution of the images [B. T. Hefner and P. L. Marston, Acoust. Res. Lett. Online **2**, 55–60 (2001)]. [Work supported by ONR.]

4:40

4pSP7. Improved ping-to-ping cross-correlation methods for synthetic aperture self-calibration. Didier Billon and Franck Fohanno (Thales Underwater Systems, B.P. 60114, 29601 Brest Cedex, France)

Ping-to-ping cross-correlation (P^2C^2) methods, like DPC (Displaced Phase Center), are an efficient way to estimate motion-related parameters for synthetic aperture beamforming, as shown by several sonar sea trials performed recently. Their accuracy depends much on the space sampling rate of the synthetic aperture. Achieving large resolution gain implies decreasing by a factor up to 2 the travel speed with respect to the absolute upper limit $L/2T$. The paper reports on two new P^2C^2 schemes that greatly relax this constraint. ''Auxiliary frequencies P^2C^2'' performs on signals from two additional frequencies transmitted at two electronically swapped phase centers close to both real aperture ends, simultaneously with the main frequency used to form the synthetic aperture. ''Hybrid P^2C^2'' consists in estimating bearing shift by means of a 2D gyrometer and an auxiliary vertical array, rather than from the main horizontal array signals. The principles of these processings are described and analytical expressions for performance bounds are given. Sonar images from a new towed sidescan sonar for deep sea floor mapping, obtained by means of both new P^2C^2 schemes, are presented. 25 cm cross-range resolution is achieved up to 800 m range. Then the resolution gain exceeds 30.

4:55

4pSP8. Extending the capabilities of synthetic aperture sonars (multiaspect and multiband) in the overall objective of buried mines classification. Alain P. Hetet, Benoit Zerr (Groupe D'Etudes Sous-Marines L'Atlantique, BP 42-29240 Brest Naval, France), and Pascal Challande (Laboratoire de Mecanique Physique, Universite Paris VI, St. Cyr, France)

Low- and medium-frequency sonars are more adequate for long-range detection and classification operations needed for mine hunting. In addition, with low frequencies the acoustic waves can penetrate into the sediment. With such a technique, detection of buried mines is possible. But, in real conditions, at sea, the number of alarms and false alarms is important. One of the problems is to find solutions to reduce those false alarms. One technique developed in this presentation is the multiband technique in association with large-bandwidth acoustical arrays. The aim of this technique is to reduce the speckle effect with noncoherent summation of multiband coherent SAS images. This technique has been focused, at the moment, on mines laying on the seabed, and is expected to be used on buried mines. The second technique discussed in the presentation is the multiaspect SAS processing applied in the low-frequency case for buried mines detection and classification. This technique is intended to help during the echo classification process of buried mines, limiting the false-alarm rate and reconstructing the shape of the object by following the echo along the shape of the objects.

THURSDAY AFTERNOON, 6 JUNE 2002 KINGS GARDEN SOUTH, 2:00 TO 4:20 P.M.

Session 4pUW

Underwater Acoustics: Shallow Water Environmental Acoustics: Session in Honor of Morris Schulkin

Robert C. Spindel, Chair

Applied Physics Laboratory, University of Washington, 1013 Northeast 40th Street, Seattle, Washington 98105-6698

Chair's Introduction—2:00

Contributed Papers

2:05

4pUW1. Comparison of ASIAEX South China Sea acoustic fluctuations with oceanographic variability. Timothy F. Duda, James F. Lynch, and Arthur E. Newhall (Appl. Ocean Phys. and Eng. Dept., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

During May 2001, signals were transmitted from five moored 16- to 100-Hz-bandwidth sources to colocated horizontal and vertical receiving arrays in 120-m-depth water on the continental slope. The site was in the northern portion of the South China Sea. Transmissions were nominally 20 km cross slope to the receiver from a site with three closely spaced sources, and 32 km up-slope to the receiver from a site with two sources, although the bathymetry was complex. Signal frequencies ranged from 210 to 550 Hz. A broad spectrum of water column variations was observed during the 17-day acoustic experiment with our array of moorings at 20 sites, including the source and receiver locations. Internal tides were consistently present, as were nonlinear internal waves of moderate amplitude. Nonlinear internal waves of larger amplitude propagating from the eastern South China Sea were present only episodically. Subtidal frequency watermass variations also occur. Correlation between various measures of the acoustic signal variability and the water column variability will be examined. [Work supported by ONR.]

2:20

4pUW2. Depth averaged acoustic field variability—its relationship to sound speed field variability—comparison of SWARM95 data to PE calculations. Peter Mignerey and Marshall Orr (Acoust. Div., The Naval Res. Lab., 4555 Overlook Ave. S.W., Washington, DC 20375)

Depth and time averaged acoustic signal energy was found to change by about 3 db over a 2.3 day period during an experiment on the New Jersey Shelf (SWARM95). The sound speed field along the 42 km propagation path has been reconstructed by integrating point CTD and range continuous acoustic flow visualization data taken during four transects of the acoustic propagation path. Sound speed fields were constructed for each transect. Depth averaged acoustic signal energy calculated for each of the sound speed fields are compared to the measured depth averaged acoustic signal energies.

2:35

4pUW3. Estimates of the prevalence of anomalous signal losses in the Yellow Sea derived from acoustic and oceanographic computer model simulations. Stanley A. Chin-Bing, David B. King, Alex C. Warn-Varnas (Naval Res. Lab., Stennis Space Center, MS 39529-5004, chinbing@nrlssc.navy.mil), Kevin G. Lamb (Univ. of Waterloo, Waterloo, ON N2L 3G1, Canada), James A. Hawkins (Planning Systems, Inc., Slidell, LA 70458), and Marvi Teixeira (Polytechnic Univ. of Puerto Rico, Hato Rey 00988, PR)

The results from collocated oceanographic and acoustic simulations in a region of the Yellow Sea near the Shandong peninsula have been presented [Chin-Bing *et al.,* J. Acoust. Soc. Am. **108**, 2577 (2000)]. In that work, the tidal flow near the peninsula was used to initialize a 2.5-dimensional ocean model [K. G. Lamb, J. Geophys. Res. **99**, 843–864 (1994)] that subsequently generated internal solitary waves (solitons). The validity of these soliton simulations was established by matching satellite imagery taken over the region. Acoustic propagation simulations through this soliton field produced results similar to the anomalous signal loss measured by Zhou, Zhang, and Rogers [J. Acoust. Soc. Am. **90**, 2042–2054 (1991)]. Analysis of the acoustic interactions with the solitons also confirmed the hypothesis that the loss mechanism involved acoustic mode coupling. Recently we have attempted to estimate the prevalence of these anomalous signal losses in this region. These estimates were made from simulating acoustic effects over an 80 hour space-time evolution of soliton packets. Examples will be presented that suggest the conditions necessary for anomalous signal loss may be more prevalent than previously thought. [Work supported by ONR/NRL and by a High Performance Computing DoD grant.]

2:50

4pUW4. Influence of environmental parameters on mid-to-high frequency acoustic signal propagation in coastal region[a] Luc Lenain, Mohsen Badiey (Ocean Acoust. Lab., College of Marine Studies, Univ. of Delaware, Newark, DE 19716), and Steve E. Forsythe (Naval Undersea Warfare Ctr., Newport, RI 02341)

Broadband mid-to-high frequency (0.6–18 kHz) acoustic wave propagation in shallow coastal waters (20 m) is influenced by a variety of oceanographic conditions. Physical parameters such as temperature and salinity as well as hydrodynamic parameters such as surface waves, tide, and current can influence amplitude and travel time of acoustic transmissions. While some of these physical parameters can be extracted from the acoustic wave propagation, understanding the forward model is vital to a fundamental understanding of the inverted parameters in shallow coastal waters. A unique set of simultaneous ocean and acoustic observations that reveal interesting temporal behavior of the acoustic signal and its correlation with environmental variability are presented. Some forward modeling using full-wave and ray-based methods is employed to investigate the time-frequency-angle characteristics of the measured acoustic field. Results of acoustic inversion for current, salinity, and sea-surface roughness, as well as the limitations of such inversions when using broadband signals in very shallow water regions, are also discussed. [Work supported by ONR and Sea Grant.] [a] For Underwater Acoustics Best Student Paper Award.

3:05–3:20 Break

3:20

4pUW5. Sound scattering in shallow water in presence of internal solitons. Boris G. Katsnelson, Serguey A. Pereselkov (Voronezh Univ., 1 Universitetskaya Sq., Voronezh 394693, Russia), and Valery G. Petnikov (Gen. Phys. Inst., 117333 Moscow, Russia)

Sound scattering by localized inhomogeneity (object) in a shallow-water waveguide is studied. Influence of internal solitons (IS) traveling in the sea, on the scattered field is considered. The problem is analyzed within the framework of theory of the sound scattering in the waveguide developed by the authors, where the scattering matrix is expressed through modal decomposition including the scattering amplitude of the object in free space. The peculiarity of the given work is the taking into account of additional modes conversion due to IS. The waveguide and IS are modeled on the basis of experimental data (sound-speed profile, bottom relief, IS parameters, etc.) collected in the Japan Sea. The inhomogeneity is selected in the form of soft spheroids with dimensions characteristic of gray whales *(eschrichtius)*. Calculations of amplitude fluctuations of the low-frequency sound field at the receiving point are carried out for moving spheroid and different orientations of acoustic track with respect to direction of propagation of IS. It is shown that sound fluctuations have essentially different characteristics for longitudinal and transversal propagation of IS. [Work supported by RFBR, Grant No. 00-05-64752.]

3:35

4pUW6. Impacts of internal tides on acoustic propagation at sonar frequencies and their effects on sonar system performance. Subramaniam D. Rajan, James K. Lewis, Jason Rudzinsky, and Peter J. Stein (Sci. Solutions, Inc., 99 Perimeter Rd., Nashua, NH 03063, srajan@scisol.com)

Temperature oscillations triggered by ocean tides can be found along nearly all shelf-slope regions throughout the world's oceans. It is of interest to ascertain how these internal tides impact acoustic propagation for operational purposes. Internal tides have been observed around the Hawaiian Islands, and these tides result in large fluctuations in temperature on a semidiurnal basis. Ocean models are presented that reproduce tidally induced sound speed variations (through model-predicted variations in temperature and salinity) around the islands of Oahu and Kauai, Hawaii. The model provides sound-speed structure and variability, and this is used as input to acoustic propagation models. Model-predicted sound-speed structures are functions of space and time. We present results of studies carried out to investigate the effect of the internal tides on the performance of sonar systems. We quantify these effects by studying the impact of internal tides on factors such as the detection range to target in active systems and source localization errors in passive systems. A means of quantifying uncertainty in the performance of the sonar systems due to the presence of the internal tides is investigated by studying the performance of the sonar system by varying the strength of the internal tidal field.

3:50

4pUW7. Challenges and possible solutions to acoustic tomography in shallow-water areas. Jason Rudzinsky, Peter J. Stein, Subramaniam D. Rajan, and James K. Lewis (Scientific Solutions, Inc., 99 Perimeter Rd., Nashua, NH 03063, jrudzinsky@scisol.com)

The Pacific Missile Range Facility off the island of Kauai in the Hawaiian chain is equipped with a large number of bottom-mounted sources and receivers. This instrumentation provides an immense opportunity for the ocean acoustic community to create an ocean acoustic laboratory. Such a laboratory can be used to monitor the changes in the three-dimensional sound-speed structure in the region using acoustic tomography. An ocean model developed for the region predicts the sound-speed structure of the region given the atmospheric, tidal, and mesoscale circulation forcing functions. This information, supplemented by acoustic tomography results, can be used to develop a consistent picture of the sound-speed field of the region. We present the challenges involved in successful implementation of tomography. The challenges are posed by the location of the sources and receivers, the bathymetry of the region, the nature of the ocean/bottom interface, the limitation on the bandwidth of the source signal, and the factors affecting the acquisition of data. We will also present approaches that can be taken to overcome some of the problems. Preliminary results of a field experiment conducted to assess the feasibility of tomography in the region will be presented.

4:05

4pUW8. Propagation of sound in the surf zone. Gary Sammelmann (Code R22 Coastal Systems Station, 6703 W. Hwy. 98, Panama City, FL 32407-7001), Grant Deane (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238), Svein Vagle, and David Farmer (Inst. of Ocean Sci., Sidney, BC V8L 4B2, Canada)

This presentation focuses on the effects of a time-varying sea surface and bubble clouds due to breaking waves on the propagation of sound in the surf zone. The rough sea surface generates caustics corresponding to Bragg diffraction due to the focusing of sound by the rough surface as described by the specular point method. A comparison with experimental data taken by Grant Deane is presented. Time evolution of bubble clouds in the surf zone is modeled by the algorithms developed by Svein Vagle, David Farmer, and Grant Deane. Bubbles in the surf zone alter both the attenuation of sound and the sound-velocity profile in the ocean. [Work supported by Codes 321OE and 32MIW of the ONR.]

Session 5aBBa

Biomedical Ultrasound/Bioresponse to Vibration: Interactions of Ultrasound with Tissue

Christy K. Holland, Cochair
Department of Radiology, University of Cincinnati, 234 Goodman Street, Cincinnati, Ohio 45219-2316

Mark E. Schafer, Cochair
Sonic Tech, Incorporated, 275 Commerce Drive, Suite 323, Fort Washington, Pennsylvania 19034

Chair's Introduction—8:00

Contributed Papers

8:05

5aBBa1. Reduction of tissue injury without compromising stone comminution in shock wave lithotripsy. Yufeng Zhou, Brian Auge, Glenn M. Preminger, and Pei Zhong (Depts. of Mech. Eng. and Mater. Sci. and Urologic Surgery, Duke Univ., Box 90300, Durham, NC 27708)

To ameliorate vascular injury without compromising stone comminution in shock wave lithotripsy, we have recently developed an *in situ* pulse superposition technique to suppress large intraluminal bubble expansion [Zhong and Zhou, J. Acoust. Soc. Am. **110**, 3283–3291 (2001)]. This strategy was implemented using a simple modification of a HM-3 lithotripter reflector. In this work, further optimization of the reflector geometry was carried out based on theoretical analysis and *in vitro* pressure waveform measurements using a fiber optical hydrophone. Using the upgraded reflector, no rupture of a cellulose hollow fiber (i.d.=0.2 mm) vessel phantom could be observed around the lithotripter beam focus even after 200 shocks at 24 kV. In comparison, less than 50 shocks were needed to cause a rupture of the vessel phantom using the original reflector at 20 kV. At corresponding output settings, stone comminution is comparable between the two reflector configurations, although the size of the fragments produced by the upgraded reflector is slightly larger. In addition, preliminary results from animal studies have demonstrated a significant reduction in tissue injury using the upgraded reflector, which confirms the validity of this approach *in vivo*. [Work supported by NIH.]

8:20

5aBBa2. Effects of an acoustic diode on lithotripter shock wave, cavitation, and stone fragmentation. Songlin Zhu (Dept. of Mech. Eng. and Mater. Sci., Duke Univ., Box 90300, Durham, NC 27708), Thomas Dreyer, Marko Liebler (Univ. of Karlsruhe, Karlsruhe, Germany), and Pei Zhong (Duke Univ., Durham, NC 27708)

Recent studies suggest that reducing the large intraluminal bubble expansion in small blood vessels may ameliorate the potential for vascular injury in shock wave lithotripsy. To achieve this objective without compromising stone comminution, a selective truncation of the tensile component of lithotripter shock wave (LSW) is needed. In this work, an acoustic diode (AD) of Riedlinger's design was constructed and evaluated. The AD consists of two peripherally secured membranes having opposite surfaces held in direct contact under partial vacuum. The AD permits transmission of the leading compressive component of a LSW; yet the membranes may separate under tension, thus blocking the transmission of the trailing tensile component of the LSW. Following each LSW, the membranes will again establish direct contact due to the partial vacuum between them. Using the AD at a vacuum level of 10.75 in. Hg, the collapse time of the LSW-induced bubble cluster at the beam focus of a HM-3 lithotripter at 20 kV was found to be reduced by 29%, whereas the compressive pressure and stone comminution were only reduced slightly by 4% and 5%, respectively. Thus the AD may be used to reduce tissue injury produced by LSW. [Work supported by Whitaker Foundation and NIH.]

8:35

5aBBa3. Cavitation bubble cluster activity in the breakage of stones by shock wave lithotripsy. Yuriy A. Pishchalnikov, Oleg A. Sapozhnikov (Dept. of Acoust., Phys. Faculty, M. V. Lomonosov Moscow State Univ., Moscow 119899, Russia, yura@acs366b.phys.msu.su), James C. Williams, Jr., Andrew P. Evan, James A. McAteer (School of Medicine, Indiana Univ., Indianapolis, IN), Robin O. Cleveland (Boston Univ., Boston, MA), Tim Colonius (California Inst. of Technol., Pasadena, CA), Michael R. Bailey, and Lawrence A. Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Seattle, WA)

High-speed photography was used to investigate cavitation at the surface of artificial and natural kidney stones during exposure to lithotripter shock pulses *in vitro*. It was observed that numerous individual bubbles formed over virtually the entire surface of the stone, but these bubbles did not remain independent and combined with one another to form larger bubbles and bubble clusters. The movement of bubble boundaries across the surface left portions of the stone bubble free. The biggest cluster grew to envelop the proximal end of the stone (6.5 mm diameter artificial stone) then collapsed to a small spot that over multiple shots formed a crater in that face of the stone. The bubble clusters that developed at the sides of stones tended to align along fractures and to collapse into these cracks. High-speed camera images demonstrated that cavitation-mediated damage to stones was due not to the action of solitary, individual bubbles, but to the forceful collapse of dynamic clusters of bubbles. [Work supported by NIH DK43881.]

8:50

5aBBa4. 30 MHz backscatter and Doppler signals from individual microbubbles undergoing inertial cavitation. Johanna M. Yoon (Eng. Dept., Swarthmore College, Swarthmore, PA 19081) and E. Carr Everbach (Swarthmore College, Swarthmore, PA 19081)

Short pulses (1–2 microseconds duration) of 30 MHz ultrasound were used to interrogate individual Optison™ and PESDA microbubbles convected in a coaxial jet flow or adhering to a thin Mylar™ window. Backscattered signals were recorded as the bubbles were forced into symmetrical or asymmetrical collapse by application of 1 MHz ultrasound pulses from a focused source in water. Peak-to-peak amplitudes of the backscattered signal were converted to radius–time curves via comparison with similar signals from a monodisperse population of polystyrene spheres of known diameter. Additionally, backscattered signals were mixed with reference sinusoids at 30 MHz and low-pass filtered to yield Doppler signals. Results are consistent with theoretical models and provide a possible method to quantify asymmetrical bubble collapse via Doppler signature.

9:05

5aBBa5. Broadband noise emissions produced by pulsed 1-MHz ultrasound exposures in the presence or absence of Optison, and their relationship to the hemolytic bioeffect. Andrew A. Brayman, Wen S. Chen, Thomas J. Matula, and Lawrence A. Crum (Ctr. for Indust. Med. Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698, brayman@apl.washington.edu)

Gas-based contrast agents are known to increase ultrasound-induced bioeffects, presumably via an inertial cavitation (IC) mechanism. The relationship between IC ''dose'' (ICD) (cumulated rms broadband noise amplitude in the frequency domain) and 1.1-MHz ultrasound-induced hemolysis in whole human blood was explored with additions of Optison or degassed saline; the hypothesis was that hemolysis would correlate with ICD. Four experimental series were conducted, with variable: (1) peak negative acoustic pressure [$P-$]; (2) Optison concentration; (3) pulse duration; and (4) total exposure duration and variable Optison concentration. The $P-$ thresholds for hemolysis and ICD above noise levels were ~ 0.5 MPa. Enhancement of ICD and hemolysis was detected even at the lowest Optison concentration tested (0.1%) at $P-=3$ MPa. At 2 MPa $P-$(0.3% Optison), significant hemolysis and ICD were detected with pulse durations as brief as 2 and 4 cycles, respectively. At 3 MPa $P-$, hemolysis and ICD evolved as functions of time and Optison concentration; ultimate levels of hemolysis and ICD depended strongly on initial Optison concentration, but initial rates of change did not. Within experimental series, hemolysis was significantly correlated with ICD; across series, the correlation was significant at p less than 0.001.

9:20

5aBBa6. Ultrasound mediated gene transfection. Rene G. Williamson, Robert E. Apfel (Dept. of Mech. Eng., Yale Univ., New Haven, CT 06520-8286, rgw27@pantheon.yale.edu), and Janet L. Brandsma (Yale Univ. School of Medicine, Cedar St., New Haven, CT 06520)

Gene therapy is a promising modality for the treatment of a variety of human diseases both inherited and acquired, such as cystic fibrosis and cancer. The lack of an effective, safe method for the delivery of foreign genes into the cells, a process known as transfection, limits this effort. Ultrasound mediated gene transfection is an attractive method for gene delivery since it is a noninvasive technique, does not introduce any viral particles into the host and can offer very good temporal and spatial control. Previous investigators have shown that sonication increases transfection efficiency with and without ultrasound contrast agents. The mechanism is believed to be via a cavitation process where collapsing bubble nuclei permeabilize the cell membrane leading to increased DNA transfer. The research is focused on the use of pulsed wave high frequency focused ultrasound to transfect DNA into mammalian cells *in vitro* and *in vivo*. A better understanding of the mechanism behind the transfection process is also sought. A summary of some *in vitro* results to date will be presented, which includes the design of a sonication chamber that allows us to model the *in vivo* case more accurately.

9:35

5aBBa7. Lung damage from exposure to low-frequency underwater sound. Diane Dalecki, Sally Z. Child, and Carol H. Raeman (Dept. of Biomed. Eng. and the Rochester Ctr. for Biomed. Ultrasound, Univ. of Rochester, Rochester, NY 14627)

The effects of low-frequency ($\sim 100-2500$ Hz) underwater sound are most pronounced in and near tissues that contain resonant gas bodies. The response of gas bodies *in vivo* (such as the lung and intestine) to low-frequency underwater sound was characterized through a series of investigations. A specially designed acoustic exposure system, capable of generating maximum acoustic fields of ~ 200 dB re: 1 microPa over the 100–2500 Hz frequency range, was implemented for these investigations. Acoustic scattering techniques were used to characterize the response of gas bodies to underwater sound exposure and to determine the resonance frequency of murine lungs. Lung damage was observed in mice exposed to underwater sound at the resonance frequency of their lung. The extent of tissue damage to the lung (and surrounding tissues such as the liver) increased with increasing pressure amplitude. Damage to lung tissue correlated with acoustic pressure amplitude and not acoustic particle velocity. Similar investigations were performed with murine intestinal gas *in vivo*.

Session 5aBBb

Biomedical Ultrasound/Bioresponse to Vibration: Scattering Theory Applications for Biomedical Media

J. Brian Fowlkes, Chair
Department of Radiology, University of Michigan Medical Center, 200 Zina Pitcher Place, Ann Arbor, Michigan 48109-0553

Chair's Introduction—10:25

Invited Papers

10:30

5aBBb1. Linking theoretical predictions of backscatter from biological media with experimental estimates. James G. Miller, Rebecca L. Trousil, Scott M. Handley, and Mark R. Holland (Dept. of Phys., Box 1105, Washington Univ., 1 Brookings Dr., Saint Louis, MO 63130)

Medical ultrasonic imaging is based on scattering processes that arise because of the inhomogeneous nature of biological media. In laboratory-based investigations, reduction of experimental rf data to the true backscatter coefficient is accomplished by compensating for measurement system, attenuation, and diffraction effects. In clinical imaging, image brightness is qualitatively related to local values of the backscatter coefficient after operator-adjusted compensation for attenuation (Time Gain Compensation). Clinical backscatter-based approaches to tissue characterization typically provide semi-quantitative data in regions-of-interest somewhat larger than the resolution cell of the image. This spatial averaging can improve the stability of the backscatter estimates. In addition, averaging over a range of frequencies can also improve the signal-to-noise ratio, as in the case of Integrated Backscatter. The sophisticated post-processing of backscattered data to form images poses an additional layer of complexity in compensating clinical data for imaging system-dependent effects. The objective of this talk is to address some approaches, and the corresponding approximations and compromises required, for comparing the results of laboratory and clinical estimates of backscatter with theoretical predictions of the backscatter coefficient (i.e., the differential scattering cross section per unit volume at 180 degrees). [Work supported in part by R37HL40302.]

11:00

5aBBb2. Statistical modeling of scattering from biological media. P. M. Shankar (ECE Dept., Drexel Univ., 3141 Chestnut St., Philadelphia, PA 19104)

The statistics of the backscattered ultrasonic echo from tissue can provide information on its characteristics. Such information is useful in the classification of tissues in biomedicine. For example, some of the tissue properties may point to malignancies in certain lesions in liver, breast, or kidneys. The models employed in describing the backscattered echo are therefore very crucial to the success of these classification methods. These models must take into account the number density of scatterers, cross sections of scatterers, variation in cross sections of the scatterers, and any alignment (periodic, quasiperiodic, and purely random) of the scatterers. Parameters reflecting these features can be extracted from the backscattered echo using these models. They can be directly related to the properties of the tissue such as the presence of an abnormal growth, and further classification of the growth as benign and malignant. They may also be used to form parametric images to assist the clinicians in making a medical diagnosis. A number of models ranging from Rayleigh, Poisson, K-, Weibull, and Nakagami will be discussed along with the relevance of their parameters and utility of the parameters in biomedicine. Specific applications to classification of breast lesions in ultrasonic B-scans will be described. [Work supported by NIH-NCI No. 52823.]

11:30

5aBBb3. High frequency ultrasonic scattering by biological tissues. K. Kirk Shung (Dept. of Bioengineering, 231 Hallowell Bldg., Penn State Univ., University Park, PA 16802) and Subha Maruvada (Brigham and Womens Hospital, Dept. of Radiol., Focused Ultrasound Group, 221 Longwood Ave., Boston, MA 02115)

High frequency (HF) diagnostic ultrasonic imaging devices at frequencies higher than 20 MHz have found applications in ophthalmology, dermatology, and vascular surgery. To be able to interpret these images and to further the development of these devices, a better understanding of ultrasonic scattering in biological tissues such as blood, liver, myocardium in the high frequency range is crucial. This work has previously been hampered by the lack of suitable transducers. With the availability of HF transducers going to 90 MHz, HF attenuation and backscatter experiments have been made on porcine red blood cell (RBC) suspensions, for which much data on attenuation and backscatter can be found in the literature in the lower frequency range for frequencies, from 30 to 90 MHz and on bovine tissues for frequencies from 10 to 30 MHz using a modified substitution method that allow the utilization of focused transducers. These results will be reviewed in this talk along with relevant theoretical models that could be applied to interpreting them. The relevance of the parameter that has been frequently used in the biomedical ultrasound literature to describe backscattering, the backscattering coefficient, will be critically examined.

12:00

5aBBb4. Multiparameter classification of masses in ultrasonic mammography. Vishruta Dumane, Mohana Shankar, Reid John (ECE Dept., Drexel Univ., 3141 Chestnut St., Philadelphia, PA 19104), Catherine Piccoli, Flemming Forsberg, and Barry Goldberg (Thomas Jefferson Univ., Philadelphia, PA 19107)

Ultrasonic characterization of breast masses for the detection of cancer can be performed based on various features of the mass. Statistical techniques that utilize the parameter of the Nakagami distribution after spatial diversity and compounding have been utilized in the past to perform the classification. The parameter demonstrated a reasonable capability for characterization of the tissue. However, there is a need to improve the performance to reach clinically acceptable standards. The work described here undertakes a combination of the Nakagami parameter after spatial compounding at the site, the skewness parameter at the site, the signal-to-noise ratio of the envelope at the site of the mass, and the margin index parameter that describes the sharpness of the boundary. The improvement in performance for characterization of the masses will be demonstrated through ROC analyses. [Work supported by NIH-NCI Grant No. CA52823.]

12:15

5aBBb5. The effect of hemolysis on acoustic scattering from blood. Constantin-C. Coussios and Shon E. Ffowcs Williams (Eng. Dept., Cambridge Univ., Trumpington St., Cambridge CB2 1PZ, UK)

In an attempt to develop a direct method for measuring the extent of red cell damage *in vitro*, the effect of the degree of hemolysis on ultrasonic scattering from blood was investigated. Starting with a suspension of 30% hematocrit, a series of suspensions containing different relative concentrations of healthy and damaged red cells in saline were prepared, with the total number of cells present in any one suspension being constant. For each sample, a suspension of equal concentration of healthy cells, but no lyzed cells, was also produced. Using a specially designed container, all samples were exposed to 15 MHz ultrasound in pulse-echo mode and measurements of backscattering were obtained. At high hematocrits, the samples containing damaged cells were found to scatter substantially more than the suspensions containing exclusively healthy cells. This indicates that damaged cells contribute significantly to the overall backscattered intensity. Below a concentration of 13% per volume of healthy cells, scattering levels from healthy and hemolyzed suspensions were comparable. A theoretical model, which treats healthy cells as weak-scattering spheres and damaged cells as hard thin disks, is proposed to interpret the observed scattering behavior.

FRIDAY MORNING, 7 JUNE 2002 STERLINGS ROOMS 2 AND 3, 8:00 TO 10:15 A.M.

Session 5aPA

Physical Acoustics: Materials Characterization, Bubbles and Drops

David B. Thiessen, Chair
Department of Physics, Washington State University, Pullman, Washington 99164-2814

Contributed Papers

8:00

5aPA1. Forgotten acousticians. Robert T. Beyer (Dept. of Phys., Brown Univ., Providence, RI 02912, rbeyer@physics.brown.edu)

In the French edition of his book on acoustics (1807), Ernest Chladni cited the work of many acousticians in or before his time. The names of Bernoulli, Biot, D'Alembert, Euler, and Laplace are familiar to us all, but Kircher, Lambert, Monro, Perolle, Riccati, Scarpa, and Mathew Young are far less so. This paper will recall the work of these individuals and others, and the contributions they made to early work on sound.

8:15

5aPA2. UV spectroscopic studies of SBSL bubbles in lithium halides. Anthony Khong, Ning Xu, Elizabeth Doschek, and Robert Apfel (Dept. of Mech. Eng., Yale Univ., 9 Hillhouse Ave., New Haven, CT 06520)

As was reported previously, stably levitated bubbles were observed in LiCl and LiBr solutions under SBSL conditions. Stable bubbles were recorded for LiCl concentrations ranging from 0.47 to 1.4 M. Beyond 1.4 M, no SL was detected. In contrast, stable SBSL can be detected over a larger range of LiBr concentrations, from 0.56 to 2.5 M. At 3.0 M, unstable short-lived transient bubbles were noticed. A striking feature common to both salt solutions is the pronounced decrease in SL light intensity, measured with a PMT with peak detection sensitivity at 400 nm, as the salt concentration increases. Light intensities were close to one order of magnitude less than in pure water under similar conditions. The focus of the current study is geared toward resolving the observed reduction in light intensity with respect to chemical processes occurring in the bubble. UV spectroscopy will be key to procuring vital information relating to these processes based on the absorption patterns and the spectral ranges of peak absorptions. Similar salt solutions in D_2O, aimed at revealing differences in the chemical processes in the case of heavy water, will also be investigated. [Work supported by a generous grant from University of Washington.]

8:30

5aPA3. Effects of mixing He and Ar on single bubble sonoluminescence (SBSL). Julio da Graça and Harry Kojima (Dept. of Phys. and Astron., Rutgers Univ., 136 Frelinghuysen Rd., Piscataway, NJ 08854-8019)

When a large pressure gradient is imposed in a gas mixture, a segregation of different species is expected to occur. Preliminary result of our search for segregation effects on the SBSL in ensonified water containing mixtures of ^{4}He (partial pressure$=p_{He}$), Ar (p_{Ar}) and N_2 (p_N) gases will be presented. Deionized and initially degassed water was mixed with $p_{He}+p_{Ar}=1.5$ Torr and $p_N=148.5$ Torr. The emitted light intensity $\mathcal{I}$ and the bubble radius dynamics (via Mie scattering) were measured as a function of the imposed 17.3 kHz acoustic drive amplitude for mixtures with varied $x=p_{He}/(p_{Ar}+p_{He})$. The acoustic pressure amplitude (P_a) at the bubble and the ambient bubble radius were extracted by fitting the Mie scattering data using Rayleigh–Plesset equation. In the range where SL occurs $[P_{lo}(x)<P_a<P_{hi}(x)]$, the measured $\mathcal{I}$ increased linearly with P_a for $x=0$, 0.5, 0.75 and 1. The slope $d\mathcal{I}/dP_a$ decreases with x. The results will be compared with expectations of homogeneous mixing of ^{4}He and Ar and segregation in the SL bubbles.

8:45

5aPA4. Optical observations of single droplets during acoustic vaporization. Oliver D. Kripfgans, Paul L. Carson, and J. Brian Fowlkes (Univ. of Michigan Health Systems, Dept. of Radiol., Ann Arbor, MI 48109, Oliver.Kripfgans@umich.edu)

It has been shown previously that Acoustic Droplet Vaporization (ADV) causes bubbles to form from small droplets. This paper shows that prior to vaporization the droplets undergo translatory (dipole) oscillations when exposed to tone bursts. A high speed video system was used to monitor droplets in a flow tube when single-element focused transducers (3.5 and 10 MHz) were used for ADV on individual droplets. Single sinusoidal tone burst (<3.3 μs) were sufficient for ADV of droplets with 5 to 25 micrometer diameter. Dipole oscillations of 1.3 μm independent of diameter were found. Variations in the droplet diameter of up to 15% were observed during the onset of acoustic irradiation. The pressure threshold for ADV decreased with increasing droplet diameter. The onset of vaporization was seen either as a localized event within or homogeneously throughout the imaged droplet (possibly due to the temporal resolution of the imaging system). Localized nucleation was solely observed along the direction of dipole motion of the droplet, which is the same as the direction of the propagating acoustic wave. Typically, sites on ''north and/or south poles'' were observed. [Research supported by PHS Grant No. R01HL54201 and US Army Grant No. DAMD17-00-1-0344.]

9:00

5aPA5. Active damping of capillary oscillations on liquid columns. David B. Thiessen, Wei Wei, and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

Active control of acoustic radiation pressure and of electrostatic stresses on liquid columns has been demonstrated to overcome the Rayleigh–Plateau instability that normally causes long liquid columns to break [M. J. Marr-Lyon *et al.,* J. Fluid Mech. **351**, 345 (1997); Phys. Fluids **12**, 986–995 (2000)]. Though originally demonstrated for liquid–liquid systems in plateau tanks, the electrostatic method also works on columns in air in reduced gravity [D. B. Thiessen, M. J. Marr-Lyon, and P. L. Marston, ''Active electrostatic stabilization of liquid bridges in low gravity,'' J. Fluid Mech. (in press)]. In new research, the electrostatic stresses are applied in proportion to the velocity of the surface of the column so as to actively dampen capillary oscillations of the surface. The mode amplitude is optically sensed and the rate-of-change is electronically determined. Plateau tank measurements and theory both show that the change in damping rate is proportional to the feedback gain. The results suggest that either active control of electrostatic stresses or of acoustic radiation stresses can be used to suppress the response of interfaces to vibration. [Work supported by NASA.]

9:15

5aPA6. Inferring pore size distributions in cast metals from ultrasonic attenuation. George Mozurkewich, Bita Ghaffari, Larry A. Godlewski, and Jacob W. Zindel (Ford Motor Co., MD3083/SRL, P.O. Box 2053, Dearborn, MI 48121-2053, gmozurke@ford.com)

The frequency dependence of ultrasonic attenuation in cast metals has previously been shown to contain information about the volume fraction and size of pores. While direct metallographic examination shows that the actual size distribution can be quite broad, previous investigations have analyzed ultrasonic attenuation data on the assumption that all pores have the same, or nearly the same, size. That restriction can be eliminated by analyzing the data using the concept of maximum entropy. Pore sizes are divided into discrete bins, each containing a fraction f_i of all the pores, and the entropy, obtained by summing $-f_i \ln(f_i)$ over all bins, is maximized subject to the constraints of the frequency-dependent attenuation data. The resulting pore size distributions for cast aluminum samples containing various levels of porosity are often found to be approximately log-normal. Volume fractions of pores deduced from these distributions are in excellent agreement with determinations using the Archimedes method.

9:30

5aPA7. Cuts with negative Poisson's ratio in alloys and rocks. Svetlana P. Tokmakova (Andreev Acoust. Inst., Shvernika 4, Moscow 117036, Russia, sveta@akin.ru)

Stretching of material with negative Poisson's ratio leads to an unexpected transverse expansion. Stretching of anisotropic material can cause expansion in one direction and contraction in another direction. For example it is possible for Poisson's ratio in crystals to be negative in one direction and highly positive in another direction. In the paper a set of hexagonal, monoclinic and cubic crystals of rocks, minerals and alloys were investigated. The stereographic projections of Poisson's ratio were computed for each crystal. From these stereographic projections the Poisson's ratio for any directions of stretch and lateral strain in crystal were calculated and orientations of stretch and lateral strain with extremes values of Poisson's ratio were obtained. By analysis of the above results cuts with negative values of Poisson's ratio were revealed in zinc, molybdenum sulfide, polypropylene, graphite, carbon, natural precious minerals labradorite and augite, complex silicate and in copper, zinc, iron, nickel, indium, silver and gold alloys. For these crystals the angular dependence of stretched orientations corresponding to negative Poisson's ratio were determined. Elastic modules of crystals from the manual of Landolt–Bornstein were applied in calculations. Some simple models with negative Poisson's ratio are considered and some possible applications of materials with negative Poisson's ratio basing on their unusual acoustic properties are discussed.

9:45

5aPA8. Acoustic measurements for nanoscale test instruments. Antanas Daugela (Hysitron, Inc., 5251 W. 73rd St., Minneapolis, MN 55439)

Acoustic methods have been used successfully for microscale contact nondestructive evaluation, but until very recently, were not available for the nanometer scale. Noncontact submicrometer resolution ultrasonic microscopy and scanning probe microscopy (SPM) operate at higher acoustic modes and have great potential to be qualitative but face data interpretation difficulties. Nanomechanical test instruments offer nanometer scale quantitative characterization. These tools coupled with *in situ* SPM-type imaging and simultaneous acoustic response monitoring open new instrumentation horizons for micro/nano fracture mechanics. Both active and passive acoustic methods can be utilized to characterize a large variety of substrates and coatings. Examples of AE monitoring of nanoindentation/scratch on data storage media will be presented. Plastic deformation induced events can be separated from contact friction events by identifying AE signatures. Active ultrasonic methods can be utilized for nanoscale characterization of tribological surfaces. A friction coefficient reduction of 20% was observed on an ultrasonically excited surface during nanoscratch testing and was further investigated using post-scratch SPM-type imaging. A synergy of localized ultrasonic monitoring and nanoindentation technique can lead to the development of new and promising instrumentation for characterizing *in vivo/vitro* biological tissues at the molecular level. Examples on evaluating ultrasonically transmitted signal through biological samples will be discussed.

10:00

5aPA9. Piezoelectric control of sculptured thin films. Fei Wang, Akhlesh Lakhtakia, and Russell Messier (Dept. of Eng. Sci. and Mech., Penn State Univ., PA 16802-6812, AXL4@psu.edu)

It shows that the Bragg center wavelength of a polymeric chiral sculptured thin film (STF) can be shifted by the axial tension generated in a co-bonded piezoelectric disk by a dc voltage. This attractive possibility can be exploited for tunable optical filters as well as lasers made of chiral STFs, and can be extended to other types of STFs.

Session 5aPPa

Psychological and Physiological Acoustics: Binaural Hearing, Monaural Phase Effects and Loudness

Douglas S. Brungart, Chair
Air Force Research Laboratory, Wright–Patterson Air Force Base, Ohio 45433-7022

Contributed Papers

8:00

5aPPa1. A counterexample of spatial unmasking in multitalker speech perception. Douglas S. Brungart (Air Force Res. Lab., 2610 Seventh St., Wright–Patterson AFB, OH 45433) and Brian D. Simpson (Veridian, 5200 Springfield Pike, Ste. 200, Dayton, OH 45431)

In listening tasks that involve more than one competing talker, substantial improvements in performance can usually be obtained by moving the target speech signal to a different location than the interfering speech signal. However, recent results in our laboratory suggest a possible listening configuration where informational and energetic masking effects might cause performance to decrease when an interfering speech signal is moved from the same ear as the target speech to the opposite ear. In this experiment, listeners were asked to respond to a color and number coordinate in a target phrase that was presented in two different listening configurations. In the first configuration, a high-level interfering talker and a low-level interfering talker were presented in the same ear as the target speech; in the second configuration, a high-level interfering talker was presented in the same ear as the target speech and a low-level interfering talker was presented in the ear opposite the target speech. The results confirm that the low-level interfering talker sometimes produced more interference when it was presented in a different ear than the target speech than when it was presented in the same ear.

8:15

5aPPa2. Transposed stimuli improve sensitivity to envelope-based interaural timing information for stimuli having center frequencies of up to 10 kHz. Leslie R. Bernstein and Constantine Trahiotis (Dept. of Neurosci. and Dept. of Surgery (Otolaryngol.), Univ. of Connecticut Health Ctr., Farmington, CT 06030, Les@neuron.uchc.ed)

Threshold interaural temporal disparities (ITDs) at high frequencies are larger than threshold ITDs obtained at low frequencies. Colburn and Esquissaud [J. Acoust. Soc. Am. Suppl. 1 **59**, S23 (1976)], hypothesized that this reflects differences in peripheral processing of the stimuli rather than in binaural mechanisms that mediate performance. Previously [L. R. Bernstein and C. Trahiotis, J. Acoust. Soc. Am. **109**, 2485 (2001)] this hypothesis was supported in ITD-discrimination experiments employing high-frequency ''transposed stimuli'' centered at 4 kHz that were designed to provide high-frequency channels with envelope-based information mimicking that normally available only in low-frequency channels. Here, we report new results using stimuli centered at 4, 6, and 10 kHz. It was found that (1) transposed stimuli can yield relatively small threshold ITDs, even at 6 and 10 kHz, (2) the data could be well accounted for in terms of a constant-criterion change in normalized interaural correlation computed subsequent to bandpass filtering, compression, rectification, and low-pass filtering. In addition, it was found necessary to incorporate a specific limitation to capture the inability of the auditory system to follow envelope fluctuations greater than 150 Hz. [Work supported by NIH DC 04147.]

8:30

5aPPa3. On the ability of human listeners to detect dispersion in head-related transfer functions. Zachary A. Constan and William M. Hartmann (Dept. of Phys. and Astron., Michigan State Univ., East Lansing, MI 48824, constan@pa.msu.edu)

Because of dispersion around the head, the interaural time difference (ITD) depends on frequency. However, virtual reality experiments have shown that human listeners cannot distinguish between veridical head-related transfer functions (HRTFs) and HRTFs with a carefully chosen constant ITD. A reasonable explanation for this result is that listeners are insensitive to the kind of dispersion created by the head. This explanation was tested in headphone experiments using incident angles and particular noise bands chosen to give listeners the best opportunity to detect dispersion. The dispersive ITD was modeled using Kuhn's equations for pressure on a spherical surface due to an incident plane wave [J. Acoust. Soc. Am. **62**, 157–167 (1977)]. Listeners were required to distinguish between noise with dispersive ITD and noise with constant ITD. Experiment 1 varied the ITD and found a pronounced minimum in the percentage of correct responses at an optimal value of ITD. Experiment 2 used optimal ITDs only in order to remove lateralization cues. Experiment 3 reversed the sign of the dispersion, causing the ITD to increase with increasing frequency rather than decrease. All experiments led to the same conclusion: listeners cannot detect head-related and similar dispersion. [Work supported by NIDCD.]

8:45

5aPPa4. Auditory feature detection for stimuli presented from different directions. Adelbert W. Bronkhorst (TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands, bronkhorst@tm.tno.nl) and James T. Townsend (Dept. of Psych., Indiana Univ., Bloomington, IN 47405)

The processing of stimulus features by the visual system is commonly studied by measuring response times (RTs) for detection of targets presented with a varying number of distracters. In audition, however, interpretation of the interaction between target and distracters is complicated by effects like masking, binaural gain and fusion. In the present paradigm, which uses stimuli presented from different directions in the horizontal plane, these confounding effects were either minimized or kept constant. Targets were synthetic vowels and RTs were measured for detection of three features: $F0$, spectrum (type of vowel) and direction. Both single features and conjunctions were studied. Independent variables were the type of distracters (synthetic vowels or spectrally shaped noise) and their number. Results show that the effect of the number of distracters on the RT is smallest for the spectrum feature; RTs for the detection of $F0$ are largest and vary considerably across subjects. For conjunctions of features, the RTs lie, as expected, between the RTs for the single features. The effect of the type of distracter on the RT is small—this indicates that the processing involved in segregating vowels from noise is not more efficient than that used for segregating vowels from other vowels.

9:00

5aPPa5. Discrimination of interaural envelope correlation for high- and low-standard correlations. Mark A. Stellmack and Neal F. Viemeister (Dept. of Psych., Univ. of Minnesota, 75 E. River Rd., Minneapolis, MN 55455)

This experiment examined the cues that listeners use in discriminating interaural envelope correlation. Thresholds for discriminating normalized interaural envelope correlation [r; see L. R. Bernstein and C. Trahiotis, J. Acoust. Soc. Am. **100**, 1754–1763 (1996)] were measured for a sinusoidally amplitude-modulated 4-kHz tone. The parameter r was manipulated by varying the interaural phase difference (IPD) of the signal envelope. Using a method of constant stimuli that measured percent correct, threshold Δr (ΔIPD) was estimated for standard $r=1$ (IPD=0 deg) and $r=0.4235$ (IPD=180 deg), with modulation frequencies (f_m) of 4–256 Hz and modulation index=0.9. For low f_m, thresholds expressed as Δr were from 10–60 times larger for the small standard correlation than for $r=1$. However, when expressed in terms of difference in the peak interaural level difference (ILD), thresholds were within several dB across standards. Furthermore, thresholds were comparable to those for discrimination of fixed ILDs. For high f_m, discrimination thresholds were smaller than those at low f_m. The difference at high f_m may be due to processing of the interaural envelope differences as fixed interaural time differences rather than dynamically varying ILDs. [Work supported by NIDCD DC00683.]

9:15

5aPPa6. Using a combined localization/detection model to simulate human localization performance near the masked detection threshold level. Jonas Braasch (Institut für Kommunikationsakustik, Ruhr-Universität Bochum, 44780 Bochum, Germany)

Recently, the perceptual lateralization of a partly masked target has been successfully simulated using the interaural cross-correlation difference (ICCD) model [Braasch, J. Acoust. Soc. Am. **108**, 2597 (2000)]. However, in the accompanying listening tests, the target level was often found to be below the masked detection threshold level. To improve the model performance, a detection threshold model has been implemented in the localization model so that the localization process is triggered only when the target level is above a threshold. Otherwise, the position of the sound source is determined according to a behavioral pattern. The detection stage is based on an on- and offset detection algorithm that analyzes the derivative in time. For the simulation of the binaural conditions, the equalization-cancellation algorithm of Durlach [J. Acoust. Soc. Am. **35**, 1206–1218 (1963)] has been included in the model. It has been shown that for target levels above the masked detection threshold, the on-/offset detection algorithm is accurate enough to trigger the subtraction process of the ICCD algorithm. Furthermore, the improved model has resulted in better simulations of human localization patterns in the presence of a distracting sound.

9:30

5aPPa7. Phase-locked onset detectors for monaural sound grouping and binaural direction finding. Leslie Smith (Dept. of Computing Sci. and Mathematics, Univ. of Stirling, Stirling FK9 4LA, Scotland, lss@cs.stir.ac.uk)

Locating sound sources is an important task for animals. IIDs and ITDs are normally used to provide information about the instantaneous direction of sound received at the ears. In a reverberating environment, this may differ from the direction of the sound source. However the IID and ITD always provide information about sound source direction at onset, since onsets always arrive from the shortest, direct path. Binaural recordings were filtered using a gammatone filterbank, converted to a phase-locked spike code, and passed to a leaky integrate-and-fire neuron through a rapidly depressing synapse. This provides a phase-locked onset detector in each bandpassed channel. Nearly coincident onsets from different channels in each ear were grouped. IIDs and ITDs were computed when grouped onsets in both ears occur at almost the same time. ITDs were converted to azimuth geometrically: IIDs were converted using the impulse response at each ear. The results show that even in a reverberating environment, sound direction can be found from a single onset. Wideband and long sounds provide better results. Multiple sound sources can be accommodated. The system exhibits the precedence effect since a second onset (without intermediate offset) will be ignored because the depressing synapses will not have recovered.

9:45

5aPPa8. Discrimination of approaching and receding sounds using three auditory motion cues. Mark Ericson (Air Force Res. Lab., AFRL/HECB Bldg. 441, Wright–Patterson AFB, OH 45433, mark.ericson@wpafb.af.mil)

The ability to determine whether a simulated moving sound source was approaching or receding was measured in several experiments using a 2 alternative, forced-choice task. Three auditory motion cues of intensity changes, Doppler frequency shifts and inter-aural time delays were encoded onto puretone stimuli and presented over headphones. Six subjects were instructed to discriminate simulated approaching versus receding sounds. The cues were presented for several velocities, durations and simulated motion paths parallel to the median-saggital and frontal planes. As found in previous studies (Ryffert *et al.*, 1979; Altman, 1999), approaching sounds encoded with monaural cues of frequency and intensity changes congruent with the frontal and median-saggital planes were detected at slower velocities than receding sounds. Subjects were more sensitive to ITD changes for paths parallel to the median-saggital plane than those parallel to the frontal plane (Mills, 1958; Hartman and Rakerd, 1985; Jenison and Wightman, 1995). Approaching versus receding discrimination thresholds were lower with the combined cues than with the three cues separated. The subjects were able to combine the monaural frequency and intensity changes and binaural inter-aural time delays to improve their judgments of approaching and receding sounds.

10:00–10:15 Break

10:15

5aPPa9. A new account of monaural phase sensitivity. Robert P. Carlyon (MRC Cognition & Brain Sci. Unit, 15 Chaucer Rd., Cambridge, UK, bob.carlyon@mrc-cbu.cam.ac.uk) and Shihab Shamma (Univ. of Maryland, College Park, MD 20742)

Phase differences between partials of complex tones are detectable only when they interact within an auditory filter. This is consistent with auditory models whereby across-channel timing information is explicitly discarded. However, these models (e.g., the autocorrelogram) fail to account for listeners' ability to detect across-channel phase differences for some sounds. For example, when two groups of unresolved harmonics are filtered into different frequency regions and presented concurrently, listeners can detect a 1- to 2-ms ''pitch pulse asynchrony'' between the peaks of the waveforms of the two groups. We attribute phase insensitivity for resolved partials to the rapid phase transition near the peak of the traveling wave, causing auditory neurons responding to each partial to do so at a variety of phases. In contrast, a group of unresolved harmonics does not produce a sharp traveling-wave peak, and all neurons responding do so synchronously. We propose a model whereby ''auditory spectrograms'' are processed by cortical filters tuned to characteristic frequency, modulation rate, and spectral ''scale.'' The model accounts quantitatively for detection of envelope phase disparities between AM tones, discrimination of AM from QFM, and discrimination of Huffman sequences. We argue that auditory models should not discard all across-channel phase information.

10:30

5aPPa10. Phase effects in simultaneous and forward masking. Elizabeth A. Lerner, Daniel L. Weber, and Brian J. Harward (Dept. of Psych., Wright State Univ., Dayton, OH 45435)

We estimated masked thresholds for a 20-ms (two 10-ms ramps), 1-kHz sinusoid as a function of its temporal relation to a 400-ms, 1-kHz sinusoid in the transition from simultaneous to forward masking condi-

tions (signal center minus masker offset times of −200, −20, −10, −3.75, −2.40, −1.25, 0.00, 1.25, 2.40, 3.75, 10, and 20 ms). In simultaneous masking conditions, thresholds for signals presented in phase with the masker were lower than for signals added in quadrature, but there was no such phase effect in forward masking conditions. When there was partial overlap of signal and masker, the phase effect was apparent until +3.75-ms center-offset time, where the phase effect disappeared and thresholds decreased. Thresholds for just the portions of the 20-ms signal that occurred during the masker (simultaneous partial signals) and for the portions of the 20-ms signals after the masker (forward partial signals), generally were consistent with results for whole (20-ms) signals: when thresholds for the whole signal appeared to be determined by simultaneous masking, thresholds for the simultaneous partial signals matched those of the whole signal and also showed a phase effect, whereas thresholds for the forward partial signals were higher and showed no phase effect.

10:45

5aPPa11. Discriminating change in the envelope phase spectrum. Stanley Sheft and William A. Yost (Parmly Hearing Inst., Loyola Univ., Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626)

Models of a modulation filterbank retain envelope phase information for only the lowest rates. To evaluate model predictions, the ability to discriminate change in the envelope phase spectrum of modulated wideband noise was measured. Modulators were narrow-band noises with bandwidth ranging from 5 to 160 Hz. The contrasting envelopes were generated by manipulating the modulator phase spectrum while leaving the amplitude spectrum unchanged. Modulator phase spectrum was varied by either reversing the modulator waveform, randomizing or zeroing phase arguments in the modulator spectrum, or generating a minimum-phase reconstruction through cepstral analysis. For time reversal and randomization, discrimination ability decreased as modulator bandwidth increased, indicating both a loss of phase information with increasing rate and masking of the low-rate information by higher-rate modulation. For zero- and minimum-phase modulators, at some point performance improved with bandwidth. This result implies a sliding temporal window to preserve power fluctuations in the filterbank output. Comparable performance was obtained when equal-bandwidth zero-phase regions were restricted to either a low- or higher-rate region. This result may reflect nonlinearity in modulation processing whereby intermodulation of high-rate components introduces perceptible low-rate modulation. [Work supported by NIH.]

11:00

5aPPa12. FM phase and the continuity illusion. Robert P. Carlyon, Christophe Micheyl, John M. Deeks (MRC Cognition & Brain Sci. Unit, 15 Chaucer Rd., Cambridge CB2 2EF, UK, bob.carlyon@mrc-cbu.cam.ac.uk), and Brian C. J. Moore (Univ. of Cambridge, Cambridge CB2 3EB, UK)

In experiment 1 subjects discriminated between a ''regular'' 1-kHz FM tone and one where the FM phase reversed midway through, at a zero crossing of the modulator waveform. Zero-peak FM depth was 100 Hz. Performance was near-perfect when the modulator frequency was 2.5 Hz, but declined sharply as it increased above 10 Hz. This is consistent with instantaneous frequency being smoothed over a finite time window, and with the phase reversal producing a ''bump'' in the window output, which decreases for faster modulations. In experiment 2 listeners discriminated between an FM tone containing a phase reversal and one where the FM depth of one-half modulator cycle was increased by various amounts (''df'')—thereby also producing a bump in the window output. Performance reached a minimum (at df= approx 5%), consistent with the theory. Finally, a 200-ms gap was inserted at the phase reversal point in a 5-Hz FM stimulus, removing subjects' ability to detect a phase reversal. Filling the gap with noise introduced a sensation of continuity, but subjects could not detect a phase reversal, even though this would have been easy in a physically continuous stimulus. Results suggest that FM phase is not encoded explicitly in the auditory system.

11:15

5aPPa13. Plasticity of loudness perception. Craig Formby, LaGuinn P. Sherlock, and Susan L. Gold (Univ. of Maryland, Baltimore, MD 21201)

Evidence from the management of tinnitus and hyperacusis suggests that loudness perception is plastic and adaptable. We have undertaken a study to evaluate this idea. The motivation followed from clinical observations suggesting that the magnitude of perceived loudness and, in turn, the rate of loudness growth can be manipulated either upward or downward by prolonged reduction or enhancement in the level of background sound to which a listener is exposed. Accordingly, volunteers were fitted bilaterally with in-the-ear noise instruments (NI treatment) or sound-attenuating earplugs (EP treatment). Both treatments produced audibility threshold shifts, mainly above 1000 Hz. The effects of each treatment were evaluated after 2 weeks of continuous use relative to pretreatment loudness response data obtained using the Contour Test of loudness perception. The resulting loudness data for warble tones revealed opposite patterns of plasticity for the two treatments, with steeper and shallower (than pretreatment) loudness growth functions measured, respectively, for the EP and NI treatments. These effects were significantly different for loudness response categories judged to be comfortably loud or louder at both 500 and 2000 Hz. Possible mechanisms for this apparent plasticity of loudness will be discussed. [Research supported by NIDCD.]

11:30

5aPPa14. Least-mean-square estimation of new equal-loudness level contours from recent data based on a loudness perception model. Yôiti Suzuki (Res. Inst. of Elec. Commun., Tohoku Univ., 2-1-1, Katahira, Aoba-ku, Sendai 980-8577, Japan) and Hisashi Takeshima (Sendai Natl. Coll. Tech., 1, Sendai 989-3124, Japan)

Since probable large errors were suggested in 1985 in the equal-loudness level contours by Robinson and Dadson [Br. J. Appl. Phys. **7**, 166–181 (1956)], which were standardized as ISO 226, a considerable amount of data on the equal-loudness relation has been accumulated. Most of the data consistently show a large discrepancy up to more than 20 phons from the contours, especially below 1 kHz. To obtain reliable contours from these new data sporadically given for some specific frequencies and phons, a model function representing the equal-loudness relation was derived from a loudness function modified by the two-stage loudness perception model. Values of the parameters of the model function were obtained by fitting the function to the experimental data. Equal-loudness level contours could be drawn by use of the model function with the parameter values interpolated along the frequency axis. The resultant equal-loudness level contours showed clear differences from those by Robinson and Dadson for all loudness levels over the whole frequency range, particularly in the frequency range lower than 1 kHz. In contrast, the contours rather resemble those given by Fletcher and Munson in 1933 and by Churcher and King in 1937 in the midfrequency range at relatively low loudness levels. [Work supported by NEDO.]

11:45

5aPPa15. Loudness for tone underwater. Edward Cudahy and Derek Schwaller (Naval Submarine Medical Res. Lab., Box 900, Groton, CT 06349-5900, e.cudahy@nsmrl.navy.mil)

The loudness for pure tones was measured by loudness matching for 1-s pure tones from 100 to 50 000 Hz. The standard tone was 1000 Hz. Subjects were instructed to match the loudness of the comparison tone at one of the test frequencies to the loudness of the standard tone. The standard was presented at one of five sound pressure levels (SPL) for each set of frequencies. The standard SPL was varied randomly across test series. The subjects were bareheaded US Navy divers tested at a depth of 3 m. All subjects had normal hearing. The tones were presented to the right side of the subject from an array of underwater sound projectors. The SPL was calibrated at the location of the subject's head with the subject absent. The loudness increased more rapidly as a function of standard SPL at midfrequencies than at either high or low frequencies. The most compact loudness contours (least SPL change across range of standard SPL) were

at 50 000 Hz. The underwater loudness contours across frequency are significantly different from in-air measurements and have a minimum in the 1000 Hz region rather than the 2–4 kHz region observed for in-air measurements. [Work supported by ONR.]

5aPPb16. Second-order temporal modulation transfer functions (TMTFs) in normal-hearing, hearing-impaired, and cochlear implant listeners. Christian Lorenzi, Jerome Sibellas (LPE, UMR CNRS 8581, Inst. de Psychologie, Univ. Paris 5, 71 Av Vaillant, 92774 Boulogne-Billancourt, France, lorenzi@psycho.univ-paris5.fr), Stephane Garnier (Groupement d'audioprothesistes ENTENDRE, 78760 Jouars-Pontchartrain, France), and Stephane Gallego (Laboratoire MXM, 06224 Vallauris Cedex, France)

''Second-order'' TMTFs are obtained by measuring detection thresholds for second-order modulation (that is, sinusoidal modulation applied to the modulation depth of a sinusoidally amplitude-modulated tone or noise carrier), as a function of fm′, the rate of second-order modulation. Here, the modulated tone or noise acts as a carrier stimulus of rate fm. Such TMTFs may be viewed as descriptions of the attenuation characteristics of the envelope beat component produced by second-order modulation at rate fm′. Second-order TMTFs assessed in normal-hearing listeners, listeners with moderately-severe cochlear damage, and cochlear implantees are similar (i.e., low pass) in shape. Overall, detection thresholds are higher in hearing-impaired than normal-hearing listeners. However, the differences in thresholds are small (5 dB). In cochlear implantees, detection thresholds are generally improved when amplitude compression is applied to the stimuli. However, the increase in sensitivity produced by compression is modest (4 dB). Taken together, these results may be accounted for in terms of modulation filters, under the following assumptions: (i) a weak distortion product is generated by the peripheral auditory system at the envelope beat rate fm′, and (ii) a salient envelope beat cue appears at the output of modulation filters tuned near the carrier rate fm.

FRIDAY MORNING, 7 JUNE 2002 GRAND BALLROOM 2, 8:30 A.M. TO 1:00 P.M.

Session 5aPPb

Psychological and Physiological Acoustics: Potpourri (Poster Session)

Christine R. Mason, Chair
Department of Communication Disorders, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

Contributed Papers

All posters will be on display from 8:30 a.m. to 1:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:45 a.m. and contributors of even-numbered papers will be at their posters from 10:45 a.m. to 1:00 p.m.

5aPPb1. Modulation rate-discrimination and frequency modulation detection thresholds in the amplitude modulation domain. Christian Füllgrabe, Christian Lorenzi (LPE-UMR CNRS 8581, Univ. Rene Descartes, Paris 5, 71 Av. Vaillant, 92774 Boulogne-Billancourt, France), and Laurent Demany (Univ. Victor Segalen, 33076 Bordeaux, France)

The present study tested the existence of modulation filters by transposing the concept of excitation pattern to the amplitude modulation (AM) domain. In the first experiment, AM rate-discrimination thresholds (DLFs) for white-noise carriers with AM reference rates of 4, 16, and 64 Hz were measured in five normal-hearing listeners. AM depth was either (i) fixed at 0.2, 0.6, or 1.0, or (ii) varied randomly between 0.2 and 1.0. Randomization of the modulation depth occurred either between trials or between the two intervals of the same trial. In a second experiment, detection thresholds of slow (1 or 4 Hz) frequency modulation (FMDLs) applied to the AM carrier (of identical rates as in the previous experiment) were measured, with and without the presence of a slow sinusoidal modulation applied to the depth of the AM carrier. Overall, the data show that DLFs are less affected than FMDLs by variations in the depth of AM. The results are discussed in light of current models of AM perception.

5aPPb2. Detection and discrimination of second-order sinusoidal amplitude modulation (SAM). Christian Füllgrabe and Christian Lorenzi (LPE-UMR CNRS 8581, Univ. Rene Descartes-Paris V, 71 Av. Vaillant, 92774 Boulogne-Billancourt, France, fullgrabe@psycho.univ-paris5.fr)

The perception of complex temporal envelopes has been recently studied using second-order SAM [Lorenzi *et al.*, J. Acoust. Soc. Am. **110**, 1030 (2001)]. In these stimuli, the modulation depth of a SAM signal (of rate fm) is sinusoidally amplitude-modulated at a rate fm', thereby generating two additional components at $fm \pm fm'$ in the modulation spectrum, and a slow beat at fm' in the temporal envelope. The present study investigates the respective contribution of these two cues to the perception of second-order SAM. In four normal-hearing listeners, second-order SAM detection and rate-discrimination abilities are measured at a high ''carrier'' ($fm = 256$ Hz), but low beat rates ($fm' \leqslant 128$ Hz), as a function of stimulus duration (250 ms to 2 s). The data are compared to first-order SAM detection and rate-discrimination thresholds measured in similar conditions at $1 \leqslant fm \leqslant 128$ Hz. At common modulation rates ($fm = fm'$), the results show that (i) first- and second-order thresholds increase similarly when stimulus duration decreases, and (ii) first- and second-order SAM rate-discrimination thresholds are basically identical. The data therefore indicate that detection and discrimination of second-order SAM are mainly based on the slow temporal envelope beat cue.

5aPPb3. The influence of practice on the detectability of auditory sinusoidal amplitude modulation. Matthew B. Fitzgerald and Beverly A. Wright (Dept. of Commun. Sci. and Disord. and the Inst. for Neurosci., Northwestern Univ., 2299 N. Campus Dr., Evanston, IL 60208-3550)

The capacity to detect fluctuations in sound amplitude influences the perception of many everyday sounds, including speech. Here, the influence of practice on this ability was investigated. Between two testing sessions, one group of nine listeners who were tested on the detection of sinusoidal amplitude modulation (SAM) improved by about 0.7 dB on each of five conditions (300 trials/condition). Nine other listeners participated in these same sessions, but between them, practiced 4320 trials

detecting the presence of 80-Hz SAM with a 3- to 4-kHz narrow-band carrier. Only three of these listeners, who had among the highest initial detection thresholds on the trained condition, improved during this training phase. These learners subsequently improved at untrained modulation rates (30 and 150 Hz) with the trained carrier, but not at the trained modulation rate with untrained carriers (0.5–1.5 kHz and 0–5 kHz). These data suggest that (1) at some stage, modulation processing is more linked to the carrier spectrum than to the modulation rate, and (2) while most normal-hearing listeners reach their best modulation-detection performance with minimal experience, listeners with high initial thresholds benefit from extended practice. Thus, training may aid populations that have difficulty detecting amplitude modulation. [Work supported by NIDCD.]

5aPPb4. Signal–masker similarity and perceptual segregation in informational masking: Some examples. Christine R. Mason, Gerald Kidd, Jr., Nathaniel I. Durlach, Tanya L. Arbogast (Hearing Res. Ctr., Boston Univ., 635 Commonwealth Ave., Boston, MA 02215, cmason@bu.edu), Barbara Shinn-Cunningham, and H. Steven Colburn (Hearing Res. Ctr., Boston Univ., 635 Commonwealth Ave., Boston, MA 02215)

In a companion paper at this meeting (Durlach *et al.*) it was suggested that changes in similarity and/or perceptual segregation of a signal and masker could affect the amount of informational masking in ways that would not be predicted based on the amount of uncertainty. In that paper, several simple stimulus manipulations are used as illustrative examples of the predicted changes. Although these were suggested as gedanken (thought) experiments, data are presented here confirming these suggestions. Perceptual segregation of the signal was accomplished by varying the degree of similarity or coherence between signal and masker along a relevant stimulus dimension. For example, when the masker consisted of a set of randomly drawn upward glides, listeners perceptually segregated a downward gliding signal. This reduced the amount of informational masking obtained relative to the condition in which both signal and masker were similar upward glides. The same was true for asynchrony of onsets, difference in perceived interaural location, and variation/coherence in spectrotemporal pattern. The purpose of these experiments was to provide data that will be useful in the development of a general theory of informational masking and multisource listening that takes into account perceptual grouping/segregation and perceived similarity. [Work supported by NIH/NIDCD.]

5aPPb5. Internal noise invariance across two informational masking tasks. Zhongzhou Tang and Virginia M. Richards (Dept. of Psych., Univ. of Pennsylvania, Philadelphia, PA 19104, richards@psych.upenn.edu)

The detectability of a 1000-Hz signal tone added to a six-tone masker was measured using a two-interval, forced-choice task. The frequencies of the masker components were randomly drawn either within a trial (different maskers in two intervals; within condition) or between trials (same maskers across two trials; between condition). Using a linear channel model with channel variances that are unaltered across conditions, weights for different frequency regions were estimated using individual responses in the within condition. Additionally, estimates of total decision variance were estimated using psychometric functions. For two observers, thresholds in the within- and between conditions were approximately the same, indicating a failure of the linear model. For a third observer the data were sufficiently variable that the psychometric functions and the linear model were poorly fitted. For the remaining five observers, thresholds in the within condition were on average 7 dB higher than in the between conditions. However, for these observers estimates of internal noise did not reliably differ across conditions. Provided the modeling assumptions are reasonable, this result suggests that for these five observers the change in threshold can be accounted for by the change in stimulus uncertainty.

5aPPb6. Sensitivity to changes in amplitude envelope. Erick Gallun, Ervin R. Hafter, and Anne-Marie Bonnel (Dept. of Psych., Univ. of California, Berkeley, CA 94720)

Detection of a brief increment in a tonal pedestal is less well predicted by energy-detection (e.g., Macmillan, 1973; Bonnel and Hafter, 1997) than by sensitivity to changes in the stimulus envelope. As this implies a mechanism similar to an envelope extractor (Viemeister, 1979), sinusoidal amplitude modulation was used to mask a single ramped increment (10, 45, or 70 ms) added to a 1000-ms pedestal with carrier frequency (cf) =477 Hz. As in informational masking (Neff, 1994) and ''modulation-detection interference'' (Yost and Sheft, 1989), interference occurred with masker cfs of 477 and 2013 Hz. While slight masking was found with modulation frequencies (mfs) from 16 to 96 Hz, masking grew inversely with still lower mfs, being greatest for mf=4 Hz. This division is reminiscent of that said to separate sensations of ''roughness'' and ''beats,'' respectively (Terhardt, 1974), with the latter also being related to durations associated with auditory groupings in music and speech. Importantly, this result held for all of the signal durations and onset–offset ramps tested, suggesting that an increment on a pedestal is treated as a single auditory object whose detection is most difficult in the presence of other objects (in this case, ''beats'').

5aPPb7. Informational masking without maskers. Robert A. Lutfi and Joshua M. Alexander (Waisman Ctr. and Dept. of Communicative Disord., Univ. of Wisconsin, Madison, WI 53706)

Informational masking is often interpreted as a failure of listeners to ''perceptually segregate'' the signal from the masker based on their different spectral-temporal properties. Though popular, such interpretations are difficult to test, as they make few specific predictions. There is, however, one prediction clearly implied by perceptual segregation. It is that informational masking should be eliminated on trials in which the masker is absent. Without a masker there can be no failure of perceptual segregation and so no masking. We report results inconsistent with this prediction. On masker-absent trials randomly interleaved with masker-present trials we show elevations in signal threshold in excess of 20 dB for some listeners. These results imply a process of perceptual summation rather than segregation. In particular, they are predicted by a model in which the decision variable is a weighted sum of signal and masker levels on each trial [Lutfi, J. Acoust. Soc. Am. **94**, 748–758 (1993)]. Elevations in signal threshold on masker-absent trials are possible according to this model because the weighted sum of levels is similar on signal-alone and masker-alone trials [Work supported by NIDCD.]

5aPPb8. Amplitude modulation perception for people with normal hearing. Jan Koopman, Niels Plasmans, and Wouter Dreschler (Dept. of Clinical & Exp. Audiol., Academic Medical Ctr., P.O. Box 22660, Meibergdreef 9 1100 DD, Amsterdam, The Netherlands, j.koopman@amc.uva.nl)

Our experiments focus on the perception of amplitude modulation. In total, four experiments were carried out. In the first experiment, the sensitivity for amplitude modulation was determined. In the other three tests, amplitude modulation has been matched for signals that differ with respect to one of the following parameters: bandwidth, center frequency, and sensation level. The equally perceived modulation depth (EPMD) was approximately equal to the reference depth for signals with a 15 dB difference in sensation level. For the bandwidth and center frequency, we generally found an interference of the internal fluctuations on the perceived modulation depth. Data can be described reasonably well by determining the differences in sensation depth (i.e., the amplitude modulation

depth [in dB] above the threshold for modulation detection). An alternative way to model the results of this study is to determine the standard deviation of energy values for a sliding temporal integrator. Both models can be improved significantly by taking into account the differences in growth of loudness functions for the two conditions.

5aPPb9. Effects of signal duration and temporal position on physiological measures of the growth of masking. Jiayun Liu and Ann Clock Eddins (Dept. of Commun. Disord. and Sci., Univ. at Buffalo, Buffalo, NY 14214)

Psychoacoustic studies have shown that the growth of masking for pure-tone signals in broadband noise is linear when the signal onset is delayed relative to the masker but is nonlinear when signal and masker onsets are synchronous [e.g., Strickland (2001)]. Curiously, for very short duration signals (10 ms), nonlinear effects have been observed for some delayed-onset conditions (Oxenham *et al.*, 1997). While nonlinear effects may result from peripheral compression, it is unclear if other physiological mechanisms might be responsible for the influence of signal temporal parameters on the growth of masking. To address this issue, the growth of masking was studied using evoked-potential responses recorded from the chinchilla inferior colliculus (IC). Quiet and masked thresholds were obtained for 1.0-, 4.0-, and 6.5-kHz signals as a function of signal duration (2–100 ms) and temporal position (delayed or nondelayed) within a 400-ms broadband noise masker. Masker spectrum levels ranged from −10 to 20 dB. Consistent with Stricklands data, physiological growth of masking was linear under delayed conditions and nonlinear under nondelayed conditions. Specifically, for nondelayed conditions, the degree of nonlinearity was greatest for high-frequency signals but did not vary with signal duration. [Work supported by NSF IBN-9996379.]

5aPPb10. Analytical predictions of auditory-nerve response to noise-modulated pulsatile electrical stimulation. William D. Ferguson, Yifang Xu (Dept. of Elec. and Computer Eng., Duke Univ., P.O. Box 90291, Durham, NC 27708, lcollins@ee.duke.edu), Roger L. Miller (Duke Univ. Medical Ctr., Durham, NC 27710), and Leslie M. Collins (Duke Univ., Durham, NC 27708)

One factor that may impede speech recognition by cochlear implant subjects is that electrically stimulated auditory nerves respond with a much higher level of synchrony than is normally observed in acoustically stimulated nerves. Thus, the response patterns received by higher processing centers are likely to be substantially different from those generated under normal acoustic stimulation. These differences may form the basis for a degradation of speech understanding since the patterns generated under electrical stimulation may be interpreted incorrectly by higher processing centers. Based on recent findings, the implant research community has suggested several techniques to mitigate this synchrony, which may in turn provide some implanted individuals with improved speech recognition. In this work, the inter-stimulus interval histogram (ISIH) is utilized to compare the response of an electrically stimulated auditory nerve with that of an acoustically stimulated nerve of a cat. The ISIH data were generated from a stochastic model of the cochlea with noise-modulated pulsatile stimulation. *Simulated* ISIHs are presented along with corroborating *analytical* predictions of the results. When compared to an acoustically generated ISIH, these data indicate that the addition of noise may provide more natural neural responses to electrical stimuli. [Work supported by NSF.]

5aPPb11. Evaluation of interval-based and beat-based timing mechanisms for duration discrimination. Melody S. Berens and Richard E. Pastore (Dept. of Psych., Binghamton Univ., PO Box 6000, Binghamton, NY 13901, mberens@binghamton.edu)

An important topic in music perception is how people track time or judge duration [E. W. Large and M. R. Jones, Psychol. Rev. **106**(1), 119–159 (1999)]. Two typically proposed duration judgment timing mechanisms are interval-based timing (judgment of discrete events) and beat-based timing (judgments of beat regularity) [S. W. Keele, N. Nicoletti, R. I. Ivry, and R. A. Pokorny, Psychol. Res. **50**, 251–256 (1989)]. Models were developed for several different versions of these hypothetical timing mechanisms. Predictions from these models were then evaluated in experiments that estimated thresholds for judging difference in judgment of interval duration. Musicianship of participants was also evaluated. Each trial began with an initial training sequence of intervals that enhanced the stored interval or established the internal rhythm or beat. The training interval was followed by a specific delay that varied across conditions. Each trial ended with a target and a comparison interval. Participants judged whether the comparison interval was equal to or longer in duration than the target interval. Results were systematic, but not consistent with simple, straightforward versions of either timing mechanism.

5aPPb12. A system for improving the communication of emotion in music performance by feedback learning. Erwin Schoonderwaldt, Anders Friberg, Roberto Bresin (Royal Inst. of Technol., Speech Music and Hearing, Drottning Kristinasv. 31, 100 44, Stockholm, Sweden, schoondw@speech.kth.se), and Patrik Juslin (Uppsala Univ., 751 42, Uppsala, Sweden)

Expressivity is one of the most important aspects of music performance. However, in music education, expressivity is often overlooked in favor of technical abilities. This could possibly depend on the difficulty in describing expressivity, which makes it problematic to provide the student with specific feedback. The aim of this project is to develop a computer program, which will improve the students' ability in communicating emotion in music performance. The expressive intention of a performer can be coded in terms of performance parameters (cues), such as tempo, sound level, timbre, and articulation. Listeners' judgments can be analyzed in the same terms. An algorithm was developed for automatic cue extraction from audio signals. Using note onset–offset detection, the algorithm yields values of sound level, articulation, IOI, and onset velocity for each note. In previous research, Juslin has developed a method for quantitative evaluation of performer–listener communication. This framework forms the basis of the present program. Multiple regression analysis on performances of the same musical fragment, played with different intentions, determines the relative importance of each cue and the consistency of cue utilization. Comparison with built-in listener models, simulating perceived expression using a regression equation, provides detailed feedback regarding the performers' cue utilization.

5aPPb13. Modeling sound transmission and reflection in the pulmonary system and chest with application to diagnosis of a collapsed lung. Thomas J. Royston, Xiangling Zhang (Univ. of Illinois at Chicago, 842 West Taylor St., MC 251, Chicago, IL 60607, troyston@uic.edu), Hussein A. Mansy, and Richard H. Sandler (Rush Medical College, Chicago, IL 60612)

Experimental studies have shown that a pneumothorax (collapsed lung) substantially alters the propagation of sound introduced at the mouth of an intubated subject and measured at the chest surface. Thus, it is hypothesized that an inexpensive diagnostic procedure could be developed for detection of a pneumothorax based on a simple acoustic test. In the present study, theoretical models of sound transmission through the pulmonary system and chest region are reviewed in the context of their ability to predict acoustic changes caused by a pneumothorax, as well as other pathologic conditions. Such models could aid in parametric design studies to develop acoustic means of diagnosing pneumothorax and other lung pathologies. Extensions of previously developed simple models of the authors are presented that are in more quantitative agreement with experimental results and that simulate both transmission from the bronchial airways to the chest wall, as well as reflection in the bronchial airways. [Research supported by NIH NCRR Grant No. 14250 and NIH NHLBI Grant No. 61108.]

5aPPb14. Differential auditory signal processing in an animal model. Dukhwan Lim, Chongsun Kim, and Sun O. Chang (Dept. of Otolaryngol., Seoul Natl. Univ., 28 Yungundong, Chongnogu, Seoul 110-744, Korea)

Auditory evoked responses were collected in male zebra finches (Poephila guttata) to objectively determine differential frequency selectivity. First, the mating call of the animal was recorded and analyzed for its frequency components through the customized program. Then, auditory brainstem responses and cortical responses of each anesthetized animal were routinely recorded in response to tone bursts of 1–8 kHz derived from the corresponding mating call spectrum. From the results, most mating calls showed relatively consistent spectral structures. The upper limit of the spectrum was well under 10 kHz. The peak energy bands were concentrated in the region less than 5 kHz. The assessment of auditory brainstem responses and cortical evoked potentials showed differential selectivity with a series of characteristic scales. This system appears to be an excellent model to investigate complex sound processing and related language behaviors. These data could also be used in designing effective signal processing strategies in auditory rehabilitation devices such as hearing aids and cochlear implants. [Work supported by Brain Science & Engineering Program from Korean Ministry of Science and Technology.]

5aPPb15. Measuring the performance of personal hearing protectors for high-noise environments. William A. Ahroon, Dale A. Ostler (U.S. Army Aeromedical Res. Lab., P.O. Box 620577, Fort Rucker, AL 36362-0577, william.ahroon@se.amedd.army.mil), and Martin B. Robinette (Lyster Army Hospital, Fort Rucker, AL 36362)

The American National Standards Institute currently identifies two real-ear attenuation at threshold (REAT) methods for measuring the performance of hearing protective devices (HPD). The experimenter-supervised-fit method (Method A), employed in earlier standards, permits the subject nearly unlimited assistance from the experimenter in fitting an HPD under test. Conversely, the subject-fit method (Method B) allows no experimenter assistance and the only fitting instructions that are available to the subject are those published by the manufacturer of the HPD and included in the HPD packaging. It is assumed that the Method A procedure provides a best-case estimate of the HPD performance while the Method B procedure provides a better estimate of the performance that can be reasonably obtained in real-world industrial settings. While measurements using the subject-fit method are preferred, there are questions as to whether this method is appropriate for measuring HPD performance in very high-noise environments such as those experienced in some military vehicles (i.e., rotary-wing aircraft and tracked vehicles). Comparisons of REAT measurements using different fitting methods are presented from studies of single- and double-protection hearing protection strategies. The preference of the ANSI S12.6-1997 subject-fit method of measuring the performance of HPDs for hearing protectors in high-level noise environments is questioned.

FRIDAY MORNING, 7 JUNE 2002 KINGS GARDEN NORTH, 8:00 TO 10:30 A.M.

Session 5aSAa

Structural Acoustics and Vibration: Modeling

Joseph M. Cuschieri, Chair
Department of Ocean Engineering, Florida Atlantic University, Center for Acoustics and Vibration, 777 Glades Road, Boca Raton, Florida 33431

Contributed Papers

8:00

5aSAa1. Dynamic stability of a beam excited by moving masses. Seroj Mackertich (The Penn State Univ., Harrisburg, Middletown, PA 17057)

The dynamic stability of an elastically supported Timoshenko beam excited by constant-velocity equally spaced traveling masses has been investigated.The regions of dynamic stability are determined for different values of the elastic foundation stiffness. Floquet theory is utilized to study the parametric regions of stability and instability, which are displayed in graphical form. Since the occurrence of this dynamic instability reduces the axial buckling load of the beam, the result is important for the study of buckling of a continuous beam.

8:15

5aSAa2. Auto-bicoherence and cross-bicoherence of the transverse vibration signals of cracked Bernoulli–Euler beams. Daniel Linehan (Appl. Res. Labs., Univ. of Texas, Austin, Austin, TX 78758)

The transverse vibration of a cracked Bernoulli–Euler beam is nonlinear if the crack does not remain open. Because this nonlinearity is even-ordered, an examination of the third-order frequency domain statistics (auto-bicoherence and cross-bicoherence) of the forcing input and acceleration output time histories should reveal quadratic nonlinearities in the form of modal interactions. To this end, the transverse vibration behavior of unconstrained, cracked Bernoulli–Euler beams was studied at the University of Texas at Austin Applied Research Laboratories. The beams were made of an aluminum 6061 alloy and were cracked with varying depths at either the midpoint or off-center. The auto-bicoherence and cross-bicoherence of both a finite-element model (FEM) and experimental beams will be discussed. The excitation was both band-limited noise and narrow-band forcing at one end of the beams. While the FEM results show a crude relationship between crack depth and modal interaction when excited with band-limited noise, the experimental results only show such a relationship when the excitation is narrow-band. The results do not demonstrate a clear method of determining the crack location. Further studies should be conducted with more discrete crack depths and different narrow-band force excitation. [Work supported by Univ. of Texas, Austin Applied Research Laboratories IR&D Program.]

8:30

5aSAa3. Mode excitation and imaging by the radiation force of ultrasound. Mostafa Fatemi and Mohammad R. Zeraati (Mayo Foundation/Clinic, Rochester, MN 55905, fatemi@mayo.edu)

A method for noncontact excitation and displaying the vibration mode shapes in solids is introduced. This method utilizes the radiation force of amplitude-modulated focused ultrasound to remotely excite a resonant mode in the object. The ultrasound transducer is driven by a sinusoidally modulated continuous wave signal to produce an oscillatory radiation force on the object and drive it into one of its natural resonance modes. The acoustic field resulting from object vibration is detected by a hydro-

phone. A theoretical model has been developed that describes the relationship between the force, mode shape, and the acoustic field. It is shown that the acoustic field amplitude is proportional to the value of the mode shape at the focal point of the ultrasound beam. By scanning the ultrasound beam across the object it is possible to map the acoustic data into an image that represents the mode shape of the object. Experiments have been conducted on small steel, aluminum, and glass beams. The first five resonant modes were excited and imaged by the present method. Experimental results have shown remarkable agreement with the theory and computer simulation. Beam deflections in the order of tens of nanometers can be detected by this method.

8:45

5aSAa4. Local/global homogenization (LGH) applied to sound reflection from a flexible barrier with impedance discontinuities. Donald Bliss, Linda Franzoni, and Pavel Danilov (Dept. of Mech. Eng., Duke Univ., Durham, NC 27708-0300)

A new homogenization method for complex structures has been developed. The method utilizes a local/global decomposition to separate the low and high parts of the wave number spectrum. The low wave number global problem has an infinite-order structural operator, and structural discontinuities are replaced by an equivalent distributed suspension. The rapidly varying local problem, which provides transfer function information for the global problem, is solved separately. Once formulated for a specific structure, the self-contained global problem is solved first, and the local solution can be reconstructed afterwards. The LGH reformulation, which applies over the entire frequency range, allows the global problem to be solved at much lower resolution than the length of flexural waves on the original structure. To demonstrate the approach, the problem of sound reflection from a flexible barrier with impedance discontinuities in a channel is described. The effects of radiating acoustic modes are transferred entirely to the smooth global problem, whereas evanescent acoustic modes are contained within the global structural operator. Sample calculations are presented comparing the method with the exact solution.

9:00

5aSAa5. Density of Axisymmetric Modes of Closed Prolate Spheriodal Shells. Courtney B. Burroughs (Appl. Research Lab., The Penn State Univ. State College, PA 16804)

The frequencies of resonance of prolate spheriodal shells may be estimated either by numerical methods or analytically via variational methods. With either approach, the computational effort increases as the order of the mode increases, making it difficult to obtain accurate estimates at frequencies where the modal density is high. However, at high frequencies, it is often necessary to obtain estimates only for the modal density since modal overlap will reduce the effects of individual modes on the response of the shell. In this paper, approximations for the modal density of the axisymmetric modes of closed prolate spheroidal shells are derived based on results obtained using analytic variational methods.

9:15–9:30 Break

9:30

5aSAa6. Structural and acoustic intensities of an infinite, point-excited fluid-loaded elastic plate. Jungyun Won and Sabih Hayek (212 EES Bldg., Penn State Univ., University Park, PA 16802-6812)

In this paper, the active vibrational structural intensity (VSI) in and the radiated acoustic intensity (AI) from an infinite elastic plate in contact with a heavy fluid is modeled by the Mindlin plate theory. This theory includes the shear deformation and rotatory inertia in addition to flexure. The plate is excited by a point force, which generates a vector active VSI field in the plate. The active VSI has two components; one depends on the shear force, and the other depends on the moment. The resulting acoustic radiation generates an active AI in the fluid medium. First, the Green's functions for the plate with and without fluid loading were developed. These were then used to develop expressions for VSI and AI vector fields. The displacement, shear deformation, VSI vector map, radiated acoustic pressure, and the AI vector map are computed for frequencies below and above the coincidence frequency. Below coincidence, a significant portion of the point force input power is trapped in the plate in the form of VSI. Above coincidence, a significant portion of the input source power is leaked to the fluid in the form of AI, with a small portion propagating to the far-field VSI.

9:45

5aSAa7. Vibration modeling of beam reinforced plates using a direct image method. Joseph Cuschieri, Alexandre Sarda (Florida Atlantic Univ., 101 N. Beach Rd., Dania, FL 33004-3023), and Arcanjo Lenzi (Universidade Federal de Santa Catarina, Florianopolis, SC, Brazil)

Plates reinforced by beams are the main components of offshore structures as used in oil prospecting and production industry. The vibration generated by machinery on these platforms propagates through the structure and can generate high noise levels in the accommodation areas. To determine the power flow through the beam reinforced plates, models that include the effect of the web and flanges' resonance in the beam and in-plane waves in all the structural components are necessary, since both of these are important especially when considering relatively high frequencies for such large structures. This work presents a model to determine the response and power flow through beam reinforced plates using a direct image method to obtain the component mobility functions for arbitrary boundary conditions. This approach is very efficient computationally and can generate accurate results up to relatively high frequencies. The results obtained using this method are compared with an exact solution using 3-D stress equations for a free rectangular parallelepiped.

10:00

5aSAa8. Experimental validation of RESOUND mid-frequency acoustic radiation models. Bryce K. Gardner and Philip J. Shorter (Vibro-Acoust. Sci., 12555 High Bluff Dr., Ste. 310, San Diego, CA 92130)

RESOUND is a full-spectrum structural acoustic analysis method that uses finite element analysis in the low frequency region, statistical energy analysis in the high frequency region and a hybrid approach in the mid frequency region. In this paper, acoustic radiation from a frame-stiffened panel into a large acoustic space will be investigated. Over the frequency range of interest, the frame has relatively few modes and exhibits long wavelength global behavior, while the panel has a large number of modes and exhibits short wavelength local behavior. Numerical and experimental results will be presented which illustrate how the frame and panel interact to give rise to the radiated sound field.

10:15

5aSAa9. Specific features of thickness resonance in finite elastic plate. Victor T. Grinchenko (Hydroacoustic Dept., Inst. of Hydromechanics of NAS of Ukraine, 8/4 Zhelyabov St., 03680 Kiev, Ukraine)

The notion of the thickness resonance in the theory of oscillation of elastic elements has been formed in the scope of the two-dimensional model for infinite elastic plate. It is presumed that the corresponding frequency is eigenfrequency for the finite plate and motion in corresponding natural mode is pistonlike. The natural modes of a finite plate are formed as a result of interaction of wave motion in plane and in thickness. Contrary to the case of ideal compressible fluid these types of motion in elastic body are tightly coupled. The interaction of two types of motion results in the essential complication of the eigenforms and spectrum of eigenfrequencies in a vicinity of the thickness resonance frequency. One of the interesting phenomena is that some eigenforms have frequencies growing together with an increase of the size of the plate. The quantitative characteristics of the eigenforms and eigenfrequency spectrum in relative high-frequency domain are presented for the case of the circular plate and finite cylinder. The distribution of displacement on the surfaces of the plate for various values of Poisson's ratio are submitted. Comparison of experimental data with results of numerical calculation is given.

Session 5aSAb

Structural Acoustics and Vibration: Radiation

John B. Fahnline, Chair
Pennsylvania State University, 16 Applied Science Building, University Park, Pennsylvania 16804

Contributed Papers

10:45

5aSAb1. Reconstruction of transient acoustic radiation from a thin disk. Manjit Bajwa and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

The HELS method [Wu, J. Acoust. Soc. Am. **107**, 2511–2522 (2000)] is extended to reconstruction of transient acoustic radiation from a highly nonspherical structure. The test object is a thin disk subject to an impulsive acceleration in an unbounded fluid medium. Since the HELS method allows piecewise reconstruction of acoustic quantities on the source surface, it is possible to focus on one side of the disk at a time. Also, since the origin of coordinates is arbitrary, one can set the spherical coordinates in such a way that the spherical surface looks almost flat locally. This treatment legitimizes the Rayleigh hypothesis and facilitates reconstruction of the normal surface velocity on the disk front surface. Reconstruction of normal surface velocity on the opposite side of the disk can be done in a similar manner. The input acoustic pressure signals are collected using an array of microphones in front of the disk and reconstructed acoustic quantities are compared with the analytic results [Wu, J. Acoust. Soc. Am. **94**, 542–553 (1993)]. Results show that the accuracy of reconstruction depends on that of input signals, and convergence of the reconstructed normal surface velocity improves with an increase in the cutoff frequency of input data. [Work supported by NSF.]

11:00

5aSAb2. Optimizing attenuation of sound radiation from a plate with passively shunted nonlinear piezoceramic patches. M. Bulent Ozer and Thomas J. Royston (Univ. of Illinois at Chicago, 842 W. Taylor St., MC 251, Chicago, IL 60607, troyston@uic.edu)

Passive control of multi-mode sound radiation from a simply supported plate using shunted piezoceramic PZT patches is investigated. Two methods are introduced and compared for calculation of the optimal inductance and resistance values of the shunt circuits: (1) the first based on adapting Den Hartogs damped vibration absorber principle, and (2) the second based on the Shermann–Morrison matrix inversion method. Both linear and nonlinear system responses are considered. At higher disturbance levels, hysteretic nonlinearity in PZT devices may degrade the passive damping performance. The modeling of this nonlinearity using an Ishlinskii hysteresis model and the determination of the optimal shunt circuit values taking hysteresis into account is discussed. Also, in this context, two different types of piezoceramic shunt circuit configurations are considered to achieve multi-mode attenuation: (1) a single PZT patch with a multiple branched shunt circuit, and (2) multiple PZT patches with single branch shunt circuits. [Research supported by NSF Grant No. 9733565 and ONR Grant No. N00014-99-1-0342.]

11:15

5aSAb3. New calculation strategy for acoustic radiation from a thick annular disk. Hyeongill Lee and Rajendra Singh (Acoust. and Dynam. Lab., Dept. of Mech. Eng. and The Ctr. for Automotive Res., The Ohio State Univ., Columbus, OH 43210-1107)

This article proposes a new semianalytical procedure for the calculation of sound radiation from a thick annular disk when it is excited by arbitrary harmonic forces. As the first step, structural eigensolutions for both in-plane (radial) and out-of-plane (flexural) modes are calculated using analytical methods. These are examined by the finite element analyses as well as the experimental investigations. The far-field sound pressure distributions (including directivity) due to selected modes of the disk are obtained from numerically obtained surface velocities using Rayleigh integral solutions based on cylindrical or circular plate radiator formulations. Such formulations define the modal radiation solutions corresponding to the structural eigensolutions of a thick disk. The boundary element analyses and vibro-acoustic experiments validate analytical predictions. Surface velocity and far-field sound pressure due to an arbitrary harmonic force excitation are then obtained from the structural and acoustic normal mode expansions. Based on the far-field sound pressure, acoustic power and radiation efficiency spectra are obtained. The method is also confirmed by comparing analytical results with those from finite and boundary element analyses. Finally, the effect of coupling between modes is also investigated by the proposed procedure.

11:30

5aSAb4. Computing multipole expansions numerically. John B. Fahnline (Appl. Res. Lab., University Park, PA 16802, jbf@wt.arl.psu.edu)

Multipole expansions have long been used to classify and better understand sound fields. They are easily computed up through the quadrupole term using standard formulas found in basic acoustic texts. Here, the formulas are adapted for numerical computations of radiated acoustic power from vibrating structures. To perform the calculations, the normal surface velocity and pressure must be known in advance and are computed using a boundary element analysis. Several examples are given to demonstrate the utility of the calculations, including a baffled circular plate clamped around its periphery and a bass-reflex loudspeaker. The results show how the expansion is useful as a way of identifying radiation mechanisms and separating out the radiating component of the surface velocity profile from the nonradiating component.

11:45

5aSAb5. Prediction of acoustic radiation based on particle velocity measurements. Qiang Hu, Zhi Ni, Huancai Lu, Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202), and Yang Zhao (Dept. of Elec. and Computer Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

It has been shown [Wu and Hu, J. Acoust. Soc. Am. **103**, 1763–1774 (1998); **104**, 3251–3258 (1998)] that the radiated acoustic pressure can be determined directly once the particle velocity distribution over an imaginary surface enclosing the object under consideration is obtained. This alternative formulation is advantageous over the classical Helmholtz integral theory, which requires the surface acoustic quantities to be completely specified before the field acoustic pressure can be calculated. The difficulty with this approach is the measurement of the fluctuating part of particle velocity in the fluid medium. This paper describes an attempt to measure particle velocities using a laser anemometer. To facilitate measurements, fine particles are sprayed in the air by a fog generator. These particles oscillate in an insonified field at the excitation frequency. Both amplitudes of particle velocities in the normal and tangential directions and phases are measured. These data are used to predict the field acoustic

pressures, which are validated by measurements taken at the same locations. Since the field acoustic pressures are calculated directly, the nonuniqueness difficulties inherent in the Helmholtz integral formulation are no longer existent and the efficiency of numerical computations is significantly enhanced. [Work supported by NSF.]

12:00

5aSAb6. An efficient method to calculate the radiated pressure from a vibrating structure. Sunghoon Choi and Yang-Hann Kim (Dept. of Mech. Eng., KAIST, Sci. Town, Taejon 305-701, Republic of Korea)

An alternative formulation of the Helmholtz integral equation, derived by Wu *et al.* [J. Acoust. Soc. Am. **103**, 1763–1774 (1998)], expresses the pressure field explicitly in terms of the velocity vector of a radiating surface. This formulation, derived for arbitrary sources, is similar in form to Rayleigh's formula for planar sources. Because the pressure field is expressed explicitly as a surface integral of the particle velocity, which can be implemented numerically using standard Gaussian quadratures, there is no need to use the boundary element method to solve a set of simultaneous equations for the surface pressure at the discretized nodes. Furthermore the nonuniqueness problem inherent in methods based on Helmholtz integral equation is avoided. Validation of this formulation is demonstrated first for some simple geometries. This method is also applied to general vibro-acoustic problems in which both the surface pressure and velocity components are unknown. [Work sponsored by Ministry of Education, Korean Government under the BK21 program and Ministry of Science and Tech., Korean Government under National Research Lab. program.]

FRIDAY MORNING, 7 JUNE 2002 GRAND BALLROOM 2, 8:30 A.M. TO 1:00 P.M.

Session 5aSC

Speech Communication: Speech Potpourri: Production and Signal Processing (Poster Session)

Fredericka Bell-Berti, Chair
Department of Speech Communication Sciences and Theatre, St. John's University, 8000 Utopia Parkway, Jamaica, New York 11439

Contributed Papers

All posters will be on display from 8:30 a.m. to 1:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:45 a.m. and contributors of even-numbered papers will be at their posters from 10:45 a.m. to 1:00 p.m.

5aSC1. Reaction time of voluntary modulations in voice *F*0 during sustained pitch vocalizations. Jay J. Bauer, Charles R. Larson (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2299 N. Campus Dr., Evanston, IL 60208), and Kathryn C. Eckstein (Univ. of Tennessee, Memphis, TN 38163)

In an attempt to more clearly understand the neural control of voice, a reaction time study was designed to investigate how rapidly normal subjects, i.e., nontrained singers, can voluntarily increase or decrease their voice fundamental frequency ($F0$) during sustained vocalizations when cued with a 1000-Hz auditory tone stimulus. Results revealed that overall reaction times (RTs) ($F=21.9$, $df=2$, 150, $p=0.01$) for upward $F0$ modulations occurred faster (range: 138–176 ms) than downward responses (range: 196–234 ms). In contrast to the reaction time findings, slightly higher peak velocities were observed for downward responses compared to upward responses. Shorter RTs observed for $F0$ elevation are therefore possibly related to central mechanisms involved in the planning of or execution of the direction in which $F0$ is to be modulated instead of muscle biomechanics. The fastest RTs obtained from the present study (138 ms) are slightly longer than the reflex latencies of the initial pitch-shift reflex response (100–130 ms) [Burnett, J. Acoust. Soc. Am. **103** (1998)], and provide additional evidence that subjects normally respond to inadvertent changes in their voice $F0$ with a fast, but limited reflex, followed by a secondary voluntary response. [Research supported by NIH Grant No. DC07264.]

5aSC2. The influence of phonotactics and phonological similarity on speech production. Michael Vitevitch, Duncan Eshelman, and Jonna Armbruster (Dept. of Psych, Univ. of Kansas, 1415 Jayhawk Blvd., Lawrence, KS 66047, mvitevit@ku.edu)

Phonotactic probability refers to the frequency with which segments and sequences of segments appear in a word or syllable. Neighborhood density refers to the number of words that are phonologically similar to a target word. These variables have been shown to influence word recognition, but little work has examined how these variables influence speech production. Although these two variables are positively correlated in English, words that varied orthogonally on these characteristics were selected and presented in a picture-naming task to assess the speed and accuracy of lexical retrieval during speech production. The results suggest that both facilitative and competitive processes operate during lexical retrieval in speech production. The implications for models of speech production are discussed. [Work funded by NIH-NIDCD R03 DC 04259.]

5aSC3. A physically informed glottis model for real glottal flow wave form reproduction. Carlo Drioli (TMH, Dept. of Speech, Music and Hearing, Royal Inst. of Technol., Drottning Kristinas v. 31, Stockholm SE 10044, Sweden, drioli@dei.unipd.it)

A physically informed model of the glottal source is proposed. The model relies on a lumped mechanoaerodynamic scheme based on the mass-spring paradigm. The vocal folds are represented by a mechanical

5a FRI. AM

resonator plus a delay line which takes into account the vertical phase differences. The vocal fold displacement is coupled to the glottal flow by means of a nonlinear subsystem, based on a general parametric nonlinear model. The principal characteristics of the flow-induced oscillations are retained, and the overall model is suited for an identification approach where real (inverse filtered) glottal flow signals are to be reproduced. A data-driven identification procedure is outlined, where the parameters of the model are tuned in order to accurately match the target wave form. A nonlinear regression algorithm is used to train the nonlinear part. A set of inverse-filtered glottal flow wave forms with different pitch, open phase/closed phase ratio, and shape of the glottal wave form period, are used to test the effectiveness of the approach. The results demonstrate that the model can reproduce a wide range of target wave forms. Moreover, the flow wave forms generated by the trained model are characterized by spectral richness and are perceived as natural. [Work supported by EU.]

5aSC4. Stress clash: Frequency and strategies of resolution. Sandra Levey (Dept. of Speech-Lang.-Hearing Sci., Lehman College, 250 Bedford Park Blvd., Bronx, NY 10468, sandralevey@hotmail.com) and Lawrence J. Raphael (Adelphi Univ., Garden City, NY 11530)

The hypothesis that speakers establish a strong–weak stress pattern to resolve stress clash (the adjacency of two primary stressed syllables, e.g., racCOON COAT) was investigated through perceptual and acoustic analysis. Three hypotheses were tested: that primary stress is relocated to an earlier syllable of the sequence (e.g., RACcoon COAT); that stress is reduced on the final syllable that bears primary stress in the first word of the sequence; that stress clash is avoided by pitch accent assignment to an early and to a late-occurring stressable syllable in a sentence. Ten speakers produced iambic target words in stress clash and non-clash contexts with target words placed in early and late-sentence position. Stress clash was resolved in less than 30% of the utterances but also occurred in non-clash contexts. Pitch accent was assigned to an early syllable for words placed in an early sentence position, increasing the frequency of stress shift judgments in both clash and non-clash contexts. Acoustic analysis showed that the information most likely to underlie shifts in stress location was located in the first syllable of target words and that fundamental frequency was the most probable of the potential cues.

5aSC5. Effects of subglottal and supraglottal acoustic loading on voice production. Zhaoyan Zhang, Luc Mongeau, and Steven Frankel (School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907)

Speech production involves sound generation by confined jets through an orifice (the glottis) with a time-varying area. Predictive models are usually based on the quasi-steady assumption. This assumption allows the complex unsteady flows to be treated as steady flows, which are more effectively modeled computationally. Because of the reflective properties of the human lungs, trachea and vocal tract, subglottal and supraglottal resonance and other acoustic effects occur in speech, which might affect glottal impedance, especially in the regime of unsteady flow separation. Changes in the flow structure, or flow regurgitation due to a transient negative transglottal pressure, could also occur. These phenomena may affect the quasi-steady behavior of speech production. To investigate the possible effects of the subglottal and supraglottal acoustic loadings, a dynamic mechanical model of the larynx was designed and built. The subglottal and supraglottal acoustic loadings are simulated using an expansion in the tube upstream of the glottis and a finite length tube downstream, respectively. The acoustic pressures of waves radiated upstream and downstream of the orifice were measured and compared to those predicted using a model based on the quasi-steady assumption. A good agreement between the experimental data and the predictions was obtained for different operating frequencies, flow rates, and orifice shapes. This supports the validity of the quasi-steady assumption for various subglottal and supraglottal acoustic loadings.

5aSC6. Kinematics of normal lingual diadokokinesis. Kevin P. Flanagan and James S. Dembowski (Commun. Disord. Dept., SUNY at New Paltz, 75 S. Manheim Blvd., New Paltz, NY 12561, dembowsj@newpaltz.edu)

Speech-language clinicians use diadochokinetic (DDK) tasks as a behavioral measure of the status of the speech production system. The articulator kinematics of these repetitive syllable productions has been relatively little studied (most studies of speech DDK have been acoustic). As a result, many clinicians misunderstand the relationship between syllable rate and movement parameters, such as rate of articulator movement and range of articulator movement. For example, clinicians assume that because ''kuh'' syllable repetitions are relatively slow, tongue dorsum movements must be slow. However, Westbury and Dembowski [Ann. Bull. RILP No. 27 (1993)] showed that the tongue dorsum may produce larger and faster movements than other tongue points (such as the tongue tip). Therefore, syllable repetition rates may not reflect movement speeds for individual articulator points. This study replicates and extends these findings, using a larger speaker sample encompassing a wider age range. The goal of this study is to further clarify the relationship between articulator speed, range of articulator motion, and syllable rate. That is, it seeks to examine whether syllable rate is related to how *quickly* the tongue moves or how *far* the tongue moves. Data are derived from the University of Wisconsin X-ray Microbeam Database.

5aSC7. Aeroacoustic mechanisms of voiced sound production. Michael Krane (CAIP Ctr., Rutgers Univ., Piscataway, NJ 08854-8088, mkrane@caip.rutgers.edu)

The focus of this study is to quantify the order of magnitude of the direct effects of (1) vocal-fold wall motion and (2) glottal flow separation point movement on the production of voiced speech sounds. A solution for the sound-pressure field shows three source mechanisms: (1) a volume source due to unsteady glottal air flow; (2) a quadrupole source representing interaction of the glottal jet with the pharynx walls; and (3) an octupole due to direct sound radiation by the glottal jet itself. A relation is derived expressing glottal volume flow in terms of transglottal pressure difference, vocal-fold wall motion, and separation point motion. Using scaling analysis, the transglottal pressure difference is shown to be the dominant effect on glottal volume flow, while vocal-fold wall motion is shown to have a negligible effect. However, separation point motion is shown to have a measurable effect during the closure phase of the vibration cycle. Using these results, the acoustic effect of separation point motion is shown to be measurable, while the effect of vocal-fold wall vibration is shown to be negligible. Relative contributions of these effects across age, gender, and degree of glottal closure are discussed.

5aSC8. Exploring the effects of gravity on tongue motion using ultrasound image sequences. Maureen Stone, Ulla Crouse, and Marty Sutton (Univ. of Maryland Dental School, Dept. of OCBS, 666 W. Baltimore St., Baltimore, MD 21201)

Our goal in the research was to explore the effect that gravity had on the vocal-tract system by using ultrasound data collected in the upright and supine positions. All potential subjects were given an ultrasound pretest to determine whether they could repeat a series of 3–4 words precise enough to allow an accurate series of images to be collected. Out of these potential subjects, approximately 5–7 subjects were eventually used in the research. The method of collecting ultrasound data required the immobilization of the patient by restraining their neck in a custom fitted neck restraint. The neck restraint held an ultrasound transducer positioned at a critical angle underneath the patients' lower jawbone, which served to

reduce errors and increase image resolution. To accurately analyze the series of images collected from ultrasound imaging, the surfaces of the tongue were digitized and tongue motion was time-aligned across the upright and supine sequences. Comparisons between the upright and supine data were then made by using L2 norms to determine averages and differences regarding the behavior between the two positions. Curves and locations of the maximum and minimum differences will be discussed.

5aSC9. Quantitative analysis of vocal fold vibration during register change by high-speed digital imaging system. Masanobu Kumada (Natl. Rehabilitation Ctr. for the Disabled, Saitama, Japan), Noriko Kobayashi, Hajime Hirose (Kitazato Univ., Kanagawa, Japan), Niro Tayama, Hiroshi Imagawa (Univ. of Tokyo, Tokyo, Japan), Ken-Ichi Sakakibara (NTT Commun. Sci. Labs., Kanagawa, Japan), Takaharu Nito, Shin'ichi Kakurai, Chieko Kumada (Univ. of Tokyo, Tokyo, Japan), Mamiko Wada (Tokyo Metropolitan Rehabilitation Hospital, Tokyo, Japan), and Seiji Niimi (Intl. Univ. for Health and Welfare, Tochigi, Japan)

The physiological study of prosody is indispensable in terms not only of the physiological interest but also of the evaluation and treatment for pathological cases of prosody. In free talk, the changes of vocal fold vibration are found frequently and these phenomena are very important prosodic events. To analyze quantitatively the vocal fold vibration at the register change as the model of prosodic event, our high-speed digital imaging system was used at a rate of 4500 images of 256−256 pixels per second. Four healthy Japanese adults (2 males and 2 females) were served as subjects. Tasks were sustained phonation containing register changes. Two major categories (Category A and B) were found in the ways of changing of vocal fold vibrations at the register change. In Category A, changes were very smooth in terms of the vocal fold vibration. In Category B, changes were not so smooth with some additional events at the register change, such as the anterior−posterior phase difference of the vibration, the abduction of the vocal folds, or the interruption of the phonation. The number of the subtypes for Category B is thought to increase if more subjects with a wider range of variety are analyzed. For the study of prosody, our high-speed digital imaging system is a very powerful tool by which physiological information can be obtained.

5aSC10. Lexical and phonotactic effects on the perception of rate induced resyllabification. Kenneth de Jong, Kyoko Nagao, Byung-jin Lim, and Kyoko Okamura (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405)

Stetson (1951) noted that, when repeated, singleton coda consonants (VC) appear to modulate into onset consonants (CV) as the rate of repetition increases. de Jong *et al*. (2001) found that naEe listeners robustly perceive such resyllabifications with labial consonants, and in a later study, that such perceptions broadly corresponded to changes in glottal timing. In the current study, stimuli included labial, coronal, and velar stops, creating mixtures of real words (such as ''eat''), and nonwords (such as ''ead''). A comparison of the perception of real and nonreal words reveals no robust effect of lexical status. In addition, vowels in the corpus were either tense or lax, so that the CV combination is phonotactically illegal in half of the corpus. The perception of resyllabification also does occur with these lax vowels, though only for voiced coronal and labial stops. Other stops did not exhibit resyllabification. Analyses of glottal and acoustic recordings are currently underway. [Work supported by NIDCD and NSF.]

5aSC11. An integrated approach to improving noisy speech perception. Serguei Koval, Mikhail Stolbov, Natalia Smirnova, and Mikhail Khitrov (STC, Krasutskogo Str. 4, St. Petersburg 196084, Russia, koval@speechpro.com)

For a number of practical purposes and tasks, experts have to decode speech recordings of very poor quality. A combination of techniques is proposed to improve intelligibility and quality of distorted speech messages and thus facilitate their comprehension. Along with the application of noise cancellation and speech signal enhancement techniques removing and/or reducing various kinds of distortions and interference (primarily unmasking and normalization in time and frequency fields), the approach incorporates optimal listener expert tactics based on selective listening, nonstandard binaural listening, accounting for short-term and long-term human ear adaptation to noisy speech, as well as some methods of speech signal enhancement to support speech decoding during listening. The approach integrating the suggested techniques ensures high-quality ultimate results and has successfully been applied by Speech Technology Center experts and by numerous other users, mainly forensic institutions, to perform noisy speech records decoding for courts, law enforcement and emergency services, accident investigation bodies, etc.

5aSC12. Stress shift in rhythmical speech. Hugo Quené and Robert Port (Dept. of Linguist., Indiana Univ., Memorial Hall 322, 1021 E. Third St., Bloomington, IN 47405)

In phrases like *thirteen men*, stress in *thirteen* is often shifted forward from its canonical final position. Presumably, the occurrence of this optional stress shift may be partly controlled by the rhythm of speech. Work on rhythmic speech production has demonstrated that given a repetition cycle, T, its harmonic fractions like $T/2$ attract stressed vowel onsets. Comparing phrases like *ceMENT thirTEEN* and *GALaxy thirTEEN*, differing in the number of weak syllables between strong ones, it was predicted that, during rhythmic production, the harmonic locations would attract shifted stress. Since shifting stress results in more even distribution of syllables through the cycle, we expected that faster repetition rates would also result in more stress shift. Dependent variables were the relative stress in the second word of each pair, and the location of onset of the nuclear vowel of the stressed syllable. Results confirmed the predictions, first, that with more intermediate unstressed syllables, stress was shifted forward more often (thereby locating the stressed vowel onset closer to $T/2$) and, second, that stress shifted forward more often at faster speaking rates. [Work supported by Fulbright Visiting Scholar program and by Utrecht University, The Netherlands.]

5aSC13. Speaker adaptation of HMMs using evolutionary strategy-based linear regression. Sid-Ahmed Selouani and Douglas O'Shaughnessy (INRS-Telecommunications, 900 de la Gauchetiere West, Box 644, Montreal, QC H5A 1C6, Canada)

A new framework for speaker adaptation of continuous-density hidden Markov models (HMMs) is introduced. It aims to improve the robustness of speech recognizers by adapting HMM parameters to new conditions (e.g., from new speakers). It describes an optimization technique using an evolutionary strategy for linear regression-based spectral transformation. In classical iterative maximum likelihood linear regression (MLLR), a global transform matrix is estimated to make a general model better match particular target conditions. To permit adaptation on a small amount of data, a regression tree classification is performed. However, an important drawback of MLLR is that the number of regression classes is fixed. The new approach allows the degree of freedom of the global transform to be implicitly variable, as the evolutionary optimization permits the survival of only active classes. The fitness function is evaluated by the phoneme correctness through the evolution steps. The implementation requirements such as chromosome representation, selection function, genetic operators, and evaluation function have been chosen in order to lend more reliability

to the global transformation matrix. Triphone experiments used the TIMIT and ARPA-RM1 databases. For new speakers, the new technique achieves 8 percent fewer word errors than the basic MLLR method.

5aSC14. The use of functional data analysis to study variability in childrens speech: Further data. Laura L. Koenig (Dept. of Speech-Lang. Pathol. and Audiol., New York Univ. and Haskins Labs., 270 Crown St., New Haven, CT 06511) and Jorge C. Lucero (Univ. of Brasilia)

Much previous research has reported increased token-to-token variability in children relative to adults, but the sources and implications of this variability remain matters of debate. Recently, functional data analysis has been used as a tool to gain greater insight into the nature of variability in children's and adults' speech data. In FDA, signals are time-normalized using a smooth function of time. The magnitude of the time-warping function provides an index of phasing (temporal) variability, and a separate index of amplitude variability is calculated from the time-normalized signal. Here, oral airflow data are analyzed from 5-year-olds, 10-year-olds, and adult women producing laryngeal and oral fricatives (/h, s, z/). The preliminary FDA results show that children generally have higher temporal and amplitude indices than adults, suggesting greater variability both in gestural timing and magnitude. However, individual patterns are evident in the relative magnitude of the two indices, and in which consonants show the highest values. The time-varying patterns of flow variability over time in /s/ are also explored as a method of inferring relative variability among laryngeal and oral gestures. [Work supported by NIH and CNPq, Brazil.]

5aSC15. Event synchronous sinusoidal model based on frequency-to-instantaneous frequency mapping. Parham Zolfaghari, Hideki Banno, Fumitada Itakura (CIAIR/CREST, Itakura Lab., Nagoya Univ., Furo-Cho 1, Chikusa-ku, Nagoya, Japan, zparham@itakura.nuee.nagoya-u.ac.jp), and Hideki Kawahara (ATR/CREST, Wakayama Univ., Japan)

We describe a glottal event synchronous sinusoidal model for speech analysis and synthesis. The sinusoidal components are event synchronously estimated using a mapping from linearly spaced filter center frequencies to the instantaneous frequencies of the filter outputs. Frequency domain fixed points of this mapping correspond to the constituent sinusoidal components of the input signal. A robust technique based on a wavelet representation of this fixed points model is used for fundamental frequency extraction as used in STRAIGHT [Kawahara *et al.,* IEICE (1999)]. The method for event detection and characterization is based on group delay and similar fixed point analysis. This method enables the detection of precise timing and spread of speech events such as vocal fold closure. A trajectory continuation scheme is also applied to the extracted sinusoidal components. The proposed model is capable of high-quality speech synthesis using the overlap–add synthesis method and is also applicable to other sound sources. System evaluation results using spectral distortion measures and mean opinion scores will be reported. A comparison with the fixed frame-rate sinusoidal models will be given.

5aSC16. Transconsonantal coarticulatory patterns in VCV utterances: Effects of a bite block. Yana Yunusova and Gary Weismer (Waisman Ctr. and Dept. of Communicative Disord., 1500 Highland Ave., Madison, WI 53705-2280)

The concept of coordinative structures, or articulatory synergies, envisions a collection of articulators organized to achieve a specific articulatory goal or acoustic goal. The mandible figures prominently in concepts of articulatory synergies because of its potential interaction with labial and lingual shaping of the vocal tract. What happens to these hypothesized synergies when one component is taken out of the collective? Bite block articulation is a common experimental approach to eliminating the jaw from its synergistic role in articulation, and studies have shown that the speech mechanism is able to reorganize its target configurations almost immediately when speaking with a fixed jaw. In the current study we examine vowel-to-vowel, transconsonantal effects with and without a bite block. Although there is one study [Sussman, Fruchter, and Cable, J. Acoust. Soc. Am. (1995)] showing stability of coarticulatory effects in biteblock conditions, that conclusion was based primarily on locus equations. In the current study, we hypothesize that more traditional acoustic measures of right-to-left and left-to-right coarticulation will show that reducing an articulatory synergy by holding one of its components constant will result in different-from-typical coarticulatory behaviors. [Work supported by NIDCD Award No. DC 000319.]

5aSC17. Dynamic synapse neural networks with a Gauss–Newton learning method for speech processing. Hassan Heidari Namarvar, Alireza Afshordi Dibazar, and Theodore W. Berger (Dept. of Biomed. Eng., Univ. of Southern California, 3650 McClintock St., OHE-500, Los Angeles, CA 90089-145, heidarin@usc.edu)

A continuous implementation of the biologically-based dynamic synapse neural network (DSNN) (H. H. Namarvar *et al.,* 2001) is created by replacing the discrete nonlinear function of the synaptic cleft mechanism, which represent the neurotransmitter release in the discrete DSNN, with a continuous nonlinear function. A Gauss–Newton learning algorithm is introduced and is shown to efficiently determine the optimal parameters of a continuous DSNN being applied to a nonlinear problem in nonstationary speech processing. The continuous DSNN incorporates a new feedback architecture to model biological inhibitory mechanisms. Optimality is determined by an objective error function on the continuous DSNN output. This network has been successfully applied to the task of phoneme recognition in continuous speech. Preliminary results demonstrate that a phoneme recognizer utilizing a continuous DSNN may be successfully used as a phone recognition module in future automatic speech recognition systems. [Work supported by DARPA CBS, NASA, and ONR.]

5aSC18. The relation between first-graders' reading level and vowel production variability and presentation format: A temporal analysis. Kandice Baker (Dept. of Speech, Commun. Sci., & Theatre, St. John's Univ., Jamaica, NY 11439), Anne Fowler (Haskins Labs., New Haven, CT), and Fredericka Bell-Berti (St. John's Univ., Jamaica, NY 11439)

The purpose of this research is to determine if children with reading difficulties produce vowels with greater variability than children with normal reading ability. The vowels chosen for this study are /ɪ/, /ɛ/, and /æ/, occurring in real and nonsense monosyllabic words. Our past research, examining spectral variability in vowels produced by first grade students as a function of whether they were reading words presented individually in random or blocked format, revealed no systematic effect of presentation format on variability [K. Baker, A. Fowler, and F. Bell-Berti, J. Acoust. Soc. Am. **110**, 2704 (2001)]. The purpose of this present study is to determine if good and poor readers differ in vowel duration variability. [Work supported by U.S. Dept. of Education, McNair Scholars Program.]

5aSC19. Applying and evaluating computer-animated tutors. Dominic W. Massaro (Dept. of Psychol., Univ. of California, Santa Cruz, Santa Cruz, CA 95064, massaro@fuzzy.ucsc.edu), Alexis Bosseler (Univ. of California, Santa Cruz, Santa Cruz, CA 95064, abosseler@yahoo.com), Patrick S. Stone (Tucker-Maxon Oral School, Portland, OR 97202, pstone@tmos.org), and Pamela Connors (Tucker-Maxon Oral School, Portland, OR 97202, connors@tmos.org)

We have developed computer-assisted speech and language tutors for deaf, hard of hearing, and autistic children. Our language-training program utilizes our computer-animated talking head, Baldi, as the conversational agent, who guides students through a variety of exercises designed to

teach vocabulary and grammer, to improve speech articulation, and to develop linguistic and phonological awareness. Baldi is an accurate three-dimensional animated talking head appropriately aligned with either synthesized or natural speech. Baldi has a tongue and palate, which can be displayed by making his skin transparent. Two specific language-training programs have been evaluated to determine if they improve word learning and speech articulation. The results indicate that the programs are effective in teaching receptive and productive language. Advantages of utilizing a computer-animated agent as a language tutor are the popularity of computers and embodied conversational agents with autistic kids, the perpetual availability of the program, and individualized instruction. Students enjoy working with Baldi because he offers extreme patience, he doesn't become angry, tired, or bored, and he is in effect a perpetual teaching machine. The results indicate that the psychology and technology of Baldi holds great promise in language learning and speech therapy. [Work supported by NSF Grant Nos. CDA-9726363 and BCS-9905176 and Public Health Service Grant No. PHS R01 DC00236.]

5aSC20. Spectral variability of /s/ in sV and sCV sequences produced by adults and children. Benjamin Munson (Dept. of Commun. Disord., Univ. of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr., SE, Minneapolis, MN 55455, Munso005@umn.edu)

Previous research has demonstrated that both children and adults produce /s/ with greater spectral variability in /sp/ sequences than /st/ sequences, when these sequences are embedded in the medial position of CVCCVC nonwords [B. Munson, J. Acoust. Soc. Am. **110**, 1203–1206 (2001)]. The current study examined whether this result could be replicated when /s/ is embedded in syllable-onset clusters, with a variety of following consonants and vowels. Adults and children aged 3–7 were recorded producing multiple tokens of sV and sCV nonwords, where the vowel was either /i/, /ɑ/, or /u/, and the consonant was either /p/, /t/, /w/, or /l/. For each token, the spectral mean of non-overlapping 10-ms windows of frication noise was calculated. Nonlinear regressions of the form $y = ae^{bx}$ were used to predict the spectral mean of each portion of frication noise from its position in the fricative. The resulting measure of model fit, R^2, was used as an index of within-speaker variability. For each participant, separate R^2 values were calculated for /s/ in each of the 15 phonetic contexts. Analyses will address the influence of age, consonant context, and vowel context on spectral variability.

5aSC21. Speech and neurology-chemical impairment correlates. Harb S. Hayre (Chemical Fitness Screening, P.O. Box 19756, Houston, TX 77224-9756)

Speech correlates of alcohol/drug impairment and its neurological basis is presented with suggestion for further research in impairment from poly drug/medicine/inhalent/chew use/abuse, and prediagnosis of many neuro- and endocrin-related disorders. Nerve cells all over the body detect chemical entry by smoking, injection, drinking, chewing, or skin absorption, and transmit neurosignals to their corresponding cerebral subsystems, which in turn affect speech centers-Broca's and Wernick's area, and motor cortex. For instance, gustatory cells in the mouth, cranial and spinal nerve cells in the skin, and cilia/olfactory neurons in the nose are the intake sensing nerve cells. Alcohol depression, and brain cell damage were detected from telephone speech using IMPAIRLYZER-TM, and the results of these studies were presented at 1996 ASA meeting in Indianapolis, and 2001 German Acoustical Society-DEGA conference in Hamburg, Germany respectively. Speech based chemical Impairment measure results were presented at the 2001 meeting of ASA in Chicago. New data on neurotolerance based chemical impairment for alcohol, drugs, and medicine shall be presented, and shown not to fully support NIDA-SAMSHA drug and alcohol threshold used in drug testing domain.

5aSC22. Voice quality variations in English sentences. Melissa Epstein (Linguist. Dept., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095-1543)

This study examines the predictability of changes in voice quality at the sentence level in English. Sentence-level effects can only be isolated once the effects of linguistic factors (e.g., glottalization before a glottalized consonant), social or dialectal, and individual factors have been eliminated. In this study, these effects were controlled by obtaining a baseline value for each measurement for each word of the corpus. Voice quality variations were tracked using quantitative measurements derived from the LF model of the glottal source, and also qualitative descriptions of the waveforms. Preliminary results indicate that there are consistent voice quality differences at the sentence level and that pitch contours and sentence accent also produce predictable effects on voice quality.

5aSC23. Characteristics of diadochokinesis in Parkinson's Disease and Multiple Sclerosis. Kris Tjaden and Elizabeth Watling (Dept. of Communicative Disord. & Sci., SUNY at Buffalo, 3435 Main St., Buffalo, NY 14214-3005)

The current study applies a quantitative, acoustic analysis procedure for the study of rapid syllable productions outlined by Kent and colleagues [R. D. Kent *et al.*, J. Med. Sp. Lang. Path. **7**, 83–90 (1999)] to syllables produced by speakers with Parkinson's Disease and Multiple Sclerosis. Neurologically healthy talkers will be studied for comparison purposes. Acoustic measures will be reported for syllable repetitions of /pə/, /tə/, and /kə/. Temporal measures will include syllable duration, syllable rate, and stop gap duration. The energy envelope of syllable repetitions will be quantified using measures of rms amplitude minima and maxima. Acoustic measures will be contrasted to determine the extent to which acoustic profiles of diadochokinesis distinguish hypokinetic dysarthria associated with Parkinson's Disease, ataxic dysarthria secondary to Multiple Sclerosis, and spastic dysarthria secondary to Multiple Sclerosis. It also is of interest to determine whether speakers with Parkinson's Disease and Multiple Sclerosis judged to be nondysarthric via perceptual analyses also demonstrate objective, acoustic profiles of diadochokinesis that are within normal limits. [Work supported by NIH.]

5aSC24. High speed MRI of laryngeal gestures during speech production. Jon Nissenbaum, Robert E. Hillman, James B. Kobler (Voice and Speech Lab., Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, MA 02114, jon_nissenbaummeei.harvard.edu), Hugh D. Curtin (Massachusetts Eye and Ear Infirmary, Boston, MA 02114), Morris Halle (MIT, Cambridge, MA 02139), and John E. Kirsch (Siemens Medical Systems, Massachusetts General Hospital, NMR Ctr., Charlestown, MA 02129)

Dynamic sequences of magnetic resonance images (MRI) of the vocal tract were obtained with a frame rate of 144 frames/second. Changes in vertical position and length of the vocal folds, both observable in the mid-sagittal plane, have been argued to play a role in consonant production in addition to their primary function in the control of vocal fundamental frequency ($F0$) [W. G. Ewan and R. Krones, J. Phonet. **2**, 327–335 (1974); A. Lofqvist *et al.*, Haskins Lab. Status Report Speech Res., SR-97/98, pp. 25–40, 1989], but temporal resolution of available techniques has hindered direct imaging of these articulations. A novel data acquisition sequence was used to circumvent the imaging time imposed by standard MRI (typically 100–500 ms). Images were constructed by having subjects rhythmically repeat short utterances 256 times using the same $F0$ contour. Sixty-four lines of MR data were sampled during each repetition, at 7 millisecond increments, yielding partial raw data sets for 64 time

points. After all repetitions were completed, one frame per time point was constructed by combining raw data from the corresponding time point during every repetition. Preliminary results indicate vocal fold shortening and lowering only during voiced consonants and in production of lower $F0$.

5aSC25. The acoustic features of human laughter. Jo-Anne Bachorowski (Dept. of Psych., Wilson Hall, Vanderbilt Univ., Nashville, TN 37203, j.a.bachorowski@vanderbilt.edu) and Michael J. Owren (Cornell Univ., Ithaca, NY 14853)

Remarkably little is known about the acoustic features of laughter, despite laughter's ubiquitous role in human vocal communication. Outcomes are described for 1024 naturally produced laugh bouts recorded from 97 young adults. Acoustic analysis focused on temporal characteristics, production modes, source- and filter-related effects, and indexical cues to laugher sex and individual identity. The results indicate that laughter is a remarkably complex vocal signal, with evident diversity in both production modes and fundamental frequency characteristics. Also of interest was finding a consistent lack of articulation effects in supralaryngeal filtering. Outcomes are compared to previously advanced hypotheses and conjectures about this species-typical vocal signal.

5aSC26. The distribution of phonation type index *k*. Hansang Park (Univ. of Texas, Cal 501, Univ. of Texas, Austin, Austin, TX 78712)

A phonation type index *k* was proposed to account for variation in the phonation type. (Park, 2001) In this study we investigated the distribution of the phonation type index *k* to see a mode of the phonation type. The distribution of the phonation type index *k* is expected to be multimodal if a phonation type is sustained through the entire vowel. However, it is expected to be unimodal if the phonation type difference is not so significant or if distinct phonation types are not maintained through the entire vowel but occur only at earlier part of the vowel. An experiment was conducted with Standard Korean data that have three linguistically distinct phonation types: aspirated, lenis, and fortis. The results showed that the distribution of the phonation type index *k* was unique to each speaker. The distribution of the phonation type index *k* was close to bimodal for one speaker while unimodal for the other two speakers. The distribution of the phonation type index *k* also showed differences in mean, standard deviation, skewness, and kurtosis across speakers.

5aSC27. Spectral moments analysis of stops in tracheoesophageal speakers. Kimberly Rosenbauer, Kerrie Obert, and Robert Allen Fox (Dept. of Speech and Hearing Sci., The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210-1002, Kimr2@aol.com)

Optimal speech intelligibility is naturally of primary concern for individuals who have had their larynxes removed due to cancer and are now using tracheoesophageal (TE) speech as their primary mode of communication. The current study examines the acoustic characteristics associated with the oral stops /p b t d k g/ produced by TE speakers as compared to normal speakers. Of particular interest are the acoustic differences between these two sets of speakers in oral stop bursts and in the aspiration frication for the voiceless stops. A set of utterances in which these six stops occur in both initial position (CV) and intervocalic position (VCV) before a wide range of English vowels were recorded for each set of speakers. Appropriate acoustic measurements were then made for each stop. These measurements included the spectral moments of the burst and aspiration, VOT, closure duration (for intervocalic stops), and the relative and normalized amplitude levels of the burst and aspiration. Acoustic differences obtained will be discussed as a function of speaker type, phonetic context and, in the case of the TE speaker, experience with the device.

5aSC28. Nonlinear viscoelastic response of vocal-fold tissues. Roger W Chan (Vocal Fold Physiol. and Biomechanics Lab., Audiol. and Speech Sci., Purdue Univ., West Lafayette, IN 47907, rchan@purdue.edu)

Previous rheological measurements of the viscoelastic shear properties of vocal-fold tissues have focused on the linear regime, in the small strain region typically with $\gamma_0 \leqslant 1.0\%$. This imposed limit was necessary in order for the theory of linear viscoelasticity to be valid, yielding dynamic shear data that can be applicable for the biomechanical modeling of small-amplitude vocal-fold oscillation. Nonetheless, as the physiological range of phonation does involve more than small-amplitude oscillation, the large strain viscoelastic behaviors of vocal-fold tissues are equally important and remain to be quantified. This paper reports preliminary measurements of some of these viscoelastic behaviors in large strain shear. Excised sheep vocal-fold mucosal tissues were subject to stress relaxation, constant stress, and constant strain rate tests in a controlled-strain torsional rheometer. Results showed that vocal-fold tissues demonstrate nonlinear viscoelastic response in shear, including stress relaxation that is dependent on strain and strain creep that is dependent on stress. These findings cannot be adequately described by Y. C. Fung's quasilinear viscoelasticity formulation, which assumes strain dependence and time dependence to be separable. A more general constitutive model is being developed to better characterize the observed nonlinear response.

5aSC29. Temporal characteristics of the speech of typical and lexically precocious two-year-old children: preliminary observations. Bruce L. Smith and Karla K. McGregor (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2299 N. Campus Dr., Evanston, IL 60208-3570, b-smith2@northwestern.edu)

To examine the extent to which temporal properties of speech might be affected by children's lexical knowledge as opposed to their age and general development, productions by a group of two-year-olds with average-sized vocabularies were compared with those of a group of age-matched, lexically precocious children. It was hypothesized that because of their additional lexical knowledge and experience, the lexically precocious children would manifest shorter durations and/or less temporal variability in their speech. Multiple repetitions of several different target words were obtained from children with vocabularies at about the 50th percentile ($ca.$ 300 words) versus the 90th percentile ($ca.$ 600 words) on the MacArthur Communicative Development Inventory. In general, acoustic measurements indicated that there were no significant differences between the groups in terms of their segmental durations or temporal variability. Thus, the additional linguistic knowledge and experience the precocious talkers had gained from having learned to produce many more words did not appear to have influenced temporal properties of their speech. This suggests that the children's age and/or other aspects of their development had a greater impact on temporal aspects of their speech than did their level of lexical knowledge and experience.

5aSC30. Harmonics to noise ratio in vocal professional voices. Luka Bonetti, Ana Bonetti, and Natalija Bolfan Stosic (Dept. of Logoped., Faculty of Special Education and Rehabilitation, Univ. of Zagreb, Kuslanova 59 a, 10000 Zagreb, Croatia)

There is no arguing about the importance of voice, especially in groups of vocal professional voices. The question is what characterizes, the most, normal or pathological voice in relation to aspects of human working life. Harmonics to noise ratio, according to findings from the field of voice disorders, is the most representative method to differ normal from pathological voice. In this research significant differences were found in harmonics to noise ratio in relation to the length of the working age of 29

teachers of primary schools in Zagreb. Teachers with the longest working age (40-yr.) showed the most distorted voices. The best quality of voice with great ratio of harmonics to noise was found in the group of teachers with 10 years of professional work. Acoustical analyses were made by EZVOICEPLUS™ version 2.0 and Gram. 2.3. Significant statistical differences were established by the T test of Statistica for Windows, version 4.5. [Work supported by Ministry of Science and Technology of Republic of Croatia.]

5aSC31. Relative timing of the three gestures of North American English /r/. Bryan Gick (Dept. of Linguist., Univ. of BC, E270-1866 Main Mall, Vancouver, BC V6T 1Z1, Canada and Haskins Labs., gick@interchange.ubc.ca) and Louis Goldstein (Yale Univ. and Haskins Labs., 270 Crown St., New Haven, CT 06511)

Interarticulator timing of liquid consonants has been shown in recent years to be of importance in understanding syllable-based allophonic variation and related synchronic, historical, and developmental phonological phenomena. In this and other respects, the unusual complexity of North American English /r/ in particular has attracted much attention. Nevertheless, because of the difficulty of collecting simultaneous dynamic measurements of the lips, tongue body and tongue root, the articulatory dynamics of /r/ remain a subject of conjecture. A simultaneous ultrasound and video study of relative timing of these three gestures of /r/ will be presented. Preliminary results from three native North American English speakers show that, for the prevocalic allophone, the lip gesture reaches its target first, followed by the tongue body (TB) gesture, then the tongue root (TR) gesture last. As for the postvocalic allophone, the lip and TR constrictions are reduced in magnitude; also the TR gesture tends to precede the other gestures in both gestural onset and peak. These results support a view of the lip gesture of /r/ as a consonantal component, and presents a pattern for /r/ analogous to that previously observed for /l/. [Research supported by NSERC and NIH.]

5aSC32. Speaker recognition using dynamic synapse-based neural networks with wavelet preprocessing. Sageev George and Theodore Berger (Dept. of Biomed. Eng., Univ. of Southern California, 3650 McClintock Ave., OHE 500, MC 1451, Los Angeles, CA 90089-1451)

Two problems in the field of speaker recognition are noise robustness and low interspeaker variability. This project involved the design of a system that is capable of speaker verification on a closed set of speakers using a wavelet processing technique that allows for a speaker-dependent feature set extraction. Verification is accomplished using a dynamic synapse-based neural network with noise-resistance properties that is trained using a genetic algorithm technique. Using these techniques, the system was able to perform speaker verification without being adversely affected by normal levels of noise, and perform verification despite low variability between speakers.

5aSC33. Talker intelligibility: Child and adult listener performance. Duncan Markham and Valerie Hazan (Dept. of Phonet. and Linguist., Univ. College, London, UK, val@phon.ucl.ac.uk)

In a study of talker intelligibility, 45 voices (adults, 11–12 year old children) were presented to 135 listeners (adults, 11–12, and 7–8 year olds). Word materials were presented in a ''single-word'' condition, and in a ''triplet'' condition, where a ''normalizing'' precursor sentence preceded three keywords. In both conditions, voices were randomized, with no consecutive presentations from the same speaker. The specially designed word-set consisted of 124 words chosen to maximize consonant confusions. Adult female speakers were significantly more intelligible than other groups, as predicted by previous research, but the difference was small. The error rates for 7–8 year olds were slightly but significantly higher than those for the older children and adults. The effect of presentation condition, however, was not significant for any listener group. Across all listener groups, rankings of speakers by error rates were strikingly consistent, with a distinct cluster of eight low-intelligibility speakers common to all listener groups. This suggests that speaker intelligibility is little influenced by listener-related factors. In terms of their perception of speaker characteristics, children aged seven and above are showing similar patterns of behavior to adults, even though the younger children showed marginally higher error rates. [Work funded by the Wellcome Trust.]

5aSC34. Adaptive interface for spoken dialog. Sorin Dusan and James Flanagan (Ctr. for Adv. Information Processing, Rutgers Univ., 96 Frelinghuysen Rd., Piscataway, NJ 08854, sdusan@caip.rutgers.edu)

Speech has become increasingly important in human–computer interaction. Spoken dialog interfaces rely on automatic speech recognition, speech synthesis, language understanding, and dialog management. A main issue in dialog systems is that they typically are limited to preprogrammed vocabularies and sets of sentences. The research reported here focuses on developing an adaptive spoken dialog interface capable of acquiring new linguistic units and their corresponding semantics during the human–computer interaction. The adaptive interface identifies unknown words and phrases in the users utterances and asks the user for the corresponding semantics. The user can provide the meaning or the semantic representation of the new linguistic units through multiple modalities, including speaking, typing, pointing, touching, or showing. The interface then stores the new linguistic units in a semantic grammar and creates new objects defining the corresponding semantic representation. This process takes place during natural interaction between user and computer and, thus, the interface does not have to be rewritten and compiled to incorporate the newly acquired language. Users can personalize the adaptive spoken interface for different domain applications, or according to their personal preferences. [Work supported by NSF.]

5aSC35. Tongue position and orientation for front vowels in the X-Ray Microbeam Speech Production DataBase. Richard S. McGowan (CReSS LLC, 1 Seaborn Pl., Lexington, MA 02420)

The positions and orientations of the secant lines between pellets will be examined for front vowels of four talkers in the X-Ray Microbeam Speech Production Database. The effect of vowel height and palate shape will be examined, as will the contextual effects of neighboring stop, fricative and nasal segments. This work is part of a project to describe tongue motion in terms of secant line kinematics. Preliminary results suggest that tongue blade orientation for high front vowels is determined largely by palate shape.

5aSC36. Isolated word recognition using dynamic synapse neural networks. Alireza A. Dibazar, Hassan H. Namarvar, and Theodore W. Berger (Dept. of Biomed. Eng., Univ. of Southern California, OHE 500, University Park, Los Angeles, CA 90089-1451, dibazar@usc.edu)

In this paper we propose a new method for using dynamic synapse neural networks (DSNNs) to accomplish isolated word recognition. The DSNNs developed by Liaw and Berger (1996) provide explicit analytic computational frameworks for the solution of nonlinear differential equations. Our method employs quasilinearization of a nonlinear differential equation to train a DSNN. This method employs an iterative algorithm, which converges monotonically to the extremal solutions of the nonlinear differential equation. The utility of the method was explored by training a simple DSNN to perform a speech recognition task on unprocessed, noisy raw waveforms of words spoken by multiple speakers. The simulation results showed that this training method has very fast convergence with respect to other existing methods. [Work supported by ONR and DARPA.]

5aSC37. Vowels in clear and conversational speech: Talker differences in acoustic characteristics and intelligibility for normal-hearing listeners. Sarah Hargus Ferguson and Diane Kewley-Port (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405)

Several studies have shown that when a talker is instructed to speak as though talking to a hearing-impaired person, the resulting ''clear'' speech is significantly more intelligible than typical conversational speech. Recent work in this lab suggests that talkers vary in how much their intelligibility improves when they are instructed to speak clearly. The few studies examining acoustic characteristics of clear and conversational speech suggest that these differing clear speech effects result from different acoustic strategies on the part of individual talkers. However, only two studies to date have directly examined differences among talkers producing clear versus conversational speech, and neither included acoustic analysis. In this project, clear and conversational speech was recorded from 41 male and female talkers aged 18–45 years. A listening experiment demonstrated that for normal-hearing listeners in noise, vowel intelligibility varied widely among the 41 talkers for both speaking styles, as did the magnitude of the speaking style effect. Acoustic analyses using stimuli from a subgroup of talkers shown to have a range of speaking style effects will be used to assess specific acoustic correlates of vowel intelligibility in clear and conversational speech. [Work supported by NIHDCD-02229.]

5aSC38. The relation between first-graders' reading level and vowel production variability in real and nonsense words: A temporal analysis. Kimberly Lydtin (Dept. of Speech, Commun. Sci., & Theatre, St. John's Univ., Jamaica, NY 11439), Anne Fowler (Haskins Labs., New Haven, CT), and Fredericka Bell-Berti (St. John's Univ., Jamaica, NY 11439)

The focus of this study is to determine if children who are poor readers produce vowels with greater variability than children with normal reading ability, since earlier research has indicated possible links between phonological difficulty, speech production variation, and reading problems. In continuation of our past research [K. Lydtin, A. Fowler, and F. Bell-Berti, J. Acoust. Soc. Am. **110**, 2704 (2001)], where we looked at the spectral aspects of vowel production, we will report the results of our study of vowel duration and its variability in poor and good readers. The vowels chosen for this study are /ɪ/, /ɛ/, and /æ/ in real and nonsense words occurring in both blocked and random presentation. [Work supported by U.S. Dept. of Education, McNair Scholars Program.]

5aSC39. Specifying voicing differences in children's productions of syllable-final stops: Knowledge versus skill. Susan Nittrouer (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

Among the acoustic correlates of phonetic identity considered to be universal is the length of vocalic segments preceding syllable-final stops, which is a correlate to the voicing of those stops. However, findings reported earlier [S. Nittrouer *et al.*, J. Acoust. Soc. Am. **109**, 2312(A) (2001)] showed that the commonly described length effect (i.e., shorter segments before voiceless than before voiced stops) is attenuated in adults' samples from continuous discourse, and that listeners of all ages fail to make much use of this effect in perceptual decisions, preferring instead to base voicing judgments on dynamic spectral information. Subsequent to that study, acoustic measures (duration of the preceding vocalic segment and frequency of the first formant at voicing offset) of children's (5 and 7 years of age) productions of words differing in the voicing of syllable-final stops showed that by 5 years of age children's productions generally had the same acoustic structure as those of adults, but within-speaker variability on both measures was roughly twice as great in children's as in adults' productions. Thus, children were trying to coordinate the vocal-tract closing and glottal abduction gestures as adults do, but were not skilled enough to do so reliably. [Work supported by NIDCD.]

FRIDAY AFTERNOON, 7 JUNE 2002 LE BATEAU ROOM, 2:00 TO 4:20 P.M.

Session 5pBB

Biomedical Ultrasound/Bioresponse to Vibration: Ultrasonic Field Characterization Techniques and Novel Instrument Applications

Thomas J. Matula, Chair

Applied Physics Laboratory, University of Washington, 1013 Northeast 40th Street, Seattle, Washington 98105-6698

Chair's Introduction—2:00

Contributed Papers

2:05

5pBB1. Wave phase conjugation of the second harmonic in a focused ultrasonic beam. A. P. Brysev, F. V. Bunkin, R. V. Klopotov, L. M. Krutyansky (Wave Res. Ctr. of the General Phys. Inst. RAS, 38 Vavilov St., 119991 Moscow GSP-1, Russia), X. Yan, and M. F. Hamilton (Univ. of Texas, Austin, TX 78712-1063)

Wave phase conjugation of the second-harmonic component generated nonlinearly in a focused beam of ultrasound is investigated experimentally and theoretically. The incident field in this case is radiated from an extended volume of the fluid between the acoustic source and the phase conjugation system. A tone burst of frequency $f=3$ MHz was radiated into water and focused at a point midway between the source and the conjugator. Phase conjugation of the second harmonic $2f$ was performed inside a magnetostrictive ceramic modulated by a magnetic pump field at frequency $4f$. The conjugate beam at frequency $2f$ reproduces quite accurately the incident second-harmonic beam everywhere between the focal plane and the conjugator. The agreement deteriorates somewhat between the focal plane and the acoustic source, because it is mainly in this region where second-harmonic generation occurs. Experimental observations are supported by analytical and numerical results. Phase conjugation using the nonlinearly generated second harmonic possesses some advantages over conventional phase conjugation of the sound beam at the source frequency. The obtained results may provide a basis for applications employ-

ing phase conjugation of harmonics in acoustic imaging and nondestructive evaluation, such as second harmonic imaging in tissue. [Work supported by RFBR, CRDF, and ONR.]

2:20

5pBB2. Mapping high power ultrasonic fields using a scanned scatterer. Bryan Cunitz and Peter Kaczkowski (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab, Univ. of Washington, Seattle, WA 98105)

The conventional method used to map the field of high intensity focused ultrasound (HIFU) transducers displaces a hydrophone over a grid of points in the zone of interest and provides a direct measure of the ultrasound pressure at each point. The approach has several major limitations: (1) the hydrophone is likely to be damaged while repeatedly measuring high intensity fields, (2) the resolution of the field map is limited to the size of the active area of the hydrophone which is typically on the order of 0.5 mm and large compared to some wavelengths of interest, and (3) cavitation can limit the accuracy of measurements. By placing a small scatterer in the HIFU field and measuring the scattered wave with a sensitive hydrophone from a safe distance, the field can be measured at full power without harm. We have used this technique to acquire single frequency field maps of HIFU transducers at high intensities without any damage to the hydrophone. We also have been able to improve the spatial resolution of the field map by an order of magnitude. In addition, this technique permits measurement of some non-linear behavior (e.g., harmonic content) at the focus at high intensities.

2:35

5pBB3. A study of angular spectrum and limited diffraction beams for calculation of field of array transducers. Jiqi Cheng and Jian-yu Lu (Ultrasound Lab., Dept. of Bioengineering, The Univ. of Toledo, Toledo, OH 43606, jilu@eng.utoledo.edu)

Angular spectrum is one of the most powerful tools for field calculation. It is based on linear system theory and the Fourier transform and is used for the calculation of propagating sound fields at different distances. In this report, the generalization and interpretation of the angular spectrum and its intrinsic relationship with limited diffraction beams are studied. With an angular spectrum, the field at the surface of a transducer is decomposed into limited diffractions beams. For an array transducer, a linear relationship between the quantized fields at the surface of elements of the array and the propagating field at any point in space can be established. For an annular array, the field is decomposed into limited diffraction Bessel beams [P. D. Fox and S. Holm, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 85–93 (2002)], while for a two-dimensional (2-D) array the field is decomposed into limited diffraction array beams [J-y. Lu and J. Cheng, J. Acoust. Soc. Am. **109**, 2397–2398 (2001)]. The angular spectrum reveals the intrinsic link between these decompositions. [Work supported in part by Grant 5RO1 HL60301 from NIH.]

2:50

5pBB4. Reconstruction of normal velocity distribution at the face of an ultrasound source in liquid on the base of acoustic waveform measurements along a surface in front of the source. Oleg A. Sapozhnikov, Yuriy A. Pishchalnikov, and Andrey V. Morozov (Dept. of Acoust., Phys. Faculty, M. V. Lomonosov Moscow State Univ., Moscow 119899, Russia, oleg@acs366b.phys.msu.su)

Normal velocity distribution along a vibrating surface is an important characteristic of any acoustic source. When it is known, the acoustic pressure field can be predicted using Rayleigh integral or similar approach. However, up to now there are no reliable methods of the velocity distribution measurement in liquids or solids. Due to strong acousto-optic interaction in condensed medium, the well-developed laser vibrometers can be employed only when the transducer is contacting vacuum or gas. In this work a novel method is developed and tested for evaluation of the velocity distribution along the vibrating surface of a piezoceramic transducer in liquid. The technique consists of measuring acoustic wave amplitude and phase along a surface surrounding the source, changing the sign of the phase, and theoretically backpropagating it to the source using the Rayleigh integral. The method was studied numerically and tested experimentally. The acoustic field of ultrasound source was registered using a needle hydrophone, which was scanned along a plane surface in front of the transducer. It is shown that the proposed approach enables accurate detection of the normal velocity. The method can be used for a wide variety of acoustically radiating structures. [Work supported by CRDF, NIH-Fogarty, and RFBR.]

3:05

5pBB5. Cyclic and radial variation of the echogenicity from human carotid artery and porcine blood. Dong-Guk Paeng, Pei-Jie Cao, K. Kirk Shung (Penn State Univ., 205 Hallowell Bldg., University Park, PA 16802, paeng@psu.edu), and Richard Y. Chiao (GE Medical Systems, EA-54, 4855 W. Electric Ave., Milwaukee, WI 53219)

The cyclic and radial variation of the echogenicity from human blood and porcine blood was investigated using a linear M12L transducer with a GE LOGIQ 700 Expert system. The bright collapsing ring phenomenon, a bright echogenic ring converging from the periphery to the center of the tube wall and eventually collapsing during a pulsatile cycle from the cross-sectional B mode images, was observed from porcine blood in a mock flow loop with a diameter of 0.95 cm at certain flow conditions. The bright ring phenomenon from porcine blood was stronger as the peak speed increased from 19 to 40 cm/s while the mean echogenicity decreased. As the stroke rate increased from 20 to 60 beats/minute, the phenomenon was weaker. As the hematocrit increased from 12 to 45%, the phenomenon became obvious. The black hole phenomenon was also observed at certain flow conditions. The well-known nonlinear hematocrit dependence on echogenicity was observed near the wall but the dependence pattern was changed at the center of the tube. The similar bright ring phenomenon was also observed at the harmonic images *in vivo* on 10 human carotid arteries. Aggregation due to the shear rate and acceleration is thought to be the explanation of the phenomena.

3:20

5pBB6. Detection of thrombosis and restenosis in an endovascular stent. Junru Wu (Dept. of Phys., Univ. of Vermont, Burlington, Burlington, VT 05405, jwu@zoo.uvm.edu) and Eric Weissman (Noveon, Inc., Brecksville, OH 44141)

Endovascular stents that are implanted in an artery are often used in the interventional treatment of coronary artery disease. Its widespread applications are, however, limited by the development of subacute thrombosis (clot forming inside of the stent). *Ex vivo* experiments with pigs have shown that the broadband A-mode ultrasound is quite effective in detection thrombosis and restenosis in an endovascular stent. [Work supported by BFGoodrich and Noveon, Inc.]

3:35

5pBB7. A robust roughness quantification technique using a standard imaging array transducer. Stanley Samuel (Univ. of Michigan Medical Ctr., 200 Zina Pitcher Pl., Rm. 3315, Kresge III, Ann Arbor, MI 48109-0553, ssamuel@umich.edu), Ronald Adler (Hospital for Special Surgery, 535 E. 70th St., New York, NY 10021), and Charles Meyer (Univ. of Michigan Medical Ctr., 200 Zina Pitcher Pl., Rm. 3315, Kresge III, Ann Arbor, MI 48109-0553)

Our goal is to measure cartilage roughness using intra-articular ultrasound imaging, thus providing a useful diagnostic tool for the early detection of osteoarthritis. Measuring the effectiveness of possible chondroprotective pharmacological or mechanical interventions depends on the availability of such a device. We have developed an empirical model of roughness using sandpaper for angles ranging from 20 degrees to 60 degrees at distances ranging from 25 mm to 80 mm. Roughness quantification is achieved using a scattering replacement normalization technique. An ultrasound imaging system employing a broadband 7 MHz multielement transducer was used for insonifying the flat sandpaper surface.

Dynamic focusing was performed at all distances and angles. The broadband transducer facilitates the selection of select bandwidths during analysis, which is beneficial for studying surfaces of varying roughness scales. Sandpaper of 150-, 400-, and 600-grit were examined for this study. The normalized average backscattered power (normalized with 150-grit) for 7–8 MHz frequency band provides well-behaved roughness characteristics. The students t test showed that the backscattering results for the 150- and 400-grit are significantly different with $0.025<p<0.05$. We are extending this work to *in vitro* cartilage and will report these results as well. [Research supported by NIH R01-AR42667-01A2.]

3:50

5pBB8. Design and construction of a high frame rate imaging system. Jing Wang, John L. Waugaman, Anjun Liu, and Jian-yu Lu (Ultrasound Lab., Dept. of Bioengineering, The Univ. of Toledo, Toledo, OH 43606, jilu@eng.utoledo.edu)

A new high frame rate imaging method has been developed recently [Jian-yu Lu, ''2D and 3D high frame rate imaging with limited diffraction beams,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 839–856 (1997)]. This method may have a clinical application for imaging of fast moving objects such as human hearts, velocity vector imaging, and low-speckle imaging. To implement the method, an imaging system has been designed. The system consists of one main printed circuit board (PCB) and 16 channel boards (each channel board contains 8 channels), in addition to a set-top box for connections to a personal computer (PC), a front panel board for user control and message display, and a power control and distribution board. The main board contains a field programmable gate array (FPGA) and controls all channels (each channel has also an FPGA). We will report the analog and digital circuit design and simulations, multiplayer PCB designs with commercial software (Protel 99), PCB signal integrity testing and system RFI/EMI shielding, and the assembly and construction of the entire system. [Work supported in part by Grant 5RO1 HL60301 from NIH.]

4:05

5pBB9. Logic design and implementation of FPGA for a high frame rate ultrasound imaging system. Anjun Liu, Jing Wang, and Jian-yu Lu (Ultrasound Lab, Dept. of Bioengineering, The Univ. of Toledo, Toledo, OH 43606, jilu@eng.utoledo.edu)

Recently, a method has been developed for high frame rate medical imaging [Jian-yu Lu, ''2D and 3D high frame rate imaging with limited diffraction beams,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**(4), 839–856 (1997)]. To realize this method, a complicated system [multiple-channel simultaneous data acquisition, large memory in each channel for storing up to 16 seconds of data at 40 MHz and 12-bit resolution, time-variable-gain (TGC) control, Doppler imaging, harmonic imaging, as well as coded transmissions] is designed. Due to the complexity of the system, field programmable gate array (FPGA) (Xilinx Spartn II) is used. In this presentation, the design and implementation of the FPGA for the system will be reported. This includes the synchronous dynamic random access memory (SDRAM) controller and other system controllers, time sharing for auto-refresh of SDRAMs to reduce peak power, transmission and imaging modality selections, ECG data acquisition and synchronization, 160 MHz delay locked loop (DLL) for accurate timing, and data transfer via either a parallel port or a PCI bus for post image processing. [Work supported in part by Grant 5RO1 HL60301 from NIH.]