CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 114, No. 3 September 2003

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 3, SEPTEMBER 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 3, SEPTEMBER 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 3, SEPTEMBER 2003

CONTENTS—*Continued from preceding page*

ACOUSTICAL NEWS—USA

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

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

The 145th meeting of the Acoustical Society of America held in Nashville, Tennessee

The 145th meeting of the Acoustical Society of America was held 28 April–2 May 2003 at the Nashville Convention Center in Nashville, Tennessee. Some events were also held at the Renaissance Nashville Hotel. This was the second time the Society has met in this city, the first time being in 1985.

The meeting drew a total of 985 registrants, including 118 nonmembers and 226 students. Attesting to the international ties of our organization, 58 of the registrants (that is, about 6%) were from outside North America. There were 19 registrants from the United Kingdom, 12 from Japan, 6 from France, 5 from Italy and the Netherlands, 3 from Korea, 2 each from Denmark and Sweden, and 1 each from Belgium, Germany, India, and Russia. North American countries, the United States, Canada, and Mexico, accounted for 899, 26, and 2, respectively.

A total of 620 papers, organized into 72 sessions, covered the areas of interest of all 13 Technical Committees. The meeting also included 13 meetings dealing with standards. The Monday evening tutorial lecture series was continued by Barbara G. Shinn-Cunningham. Her tutorial ''Hearing in Three Dimensions'' was presented to an audience of about 115 people.

A social hour was held on Wednesday evening at The Ryman Auditorium which served as the home of the Grand Ole Opry from 1943 to 1974. It is now a museum and a performance hall. The local committee arranged for a live performance by Marcus Hummon during the social hour. Other events included a social hour held at the Renaissance Hotel on Tuesday, receptions for students in various areas of acoustics, a Fellows Lounge, and 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.

There were several special technical events held at the meeting. A ''hot topics'' session included presentations on Biomedical Ultrasound/Bioresponse to Vibration, Education in Acoustics, and Underwater Acoustics. The Acoustical Oceanography Mini-Tutorial Series, with the goal of providing up-to-date overviews of ocean processes that affect acoustics and that in turn can be studied using acoustics methods, included two presentations. The Technical Committee on Architectural Acoustics and the National Council of Acoustical Consultants sponsored a student design competition which was professionally judged at the meeting. The winners were: First Honors ($1000 prize): Brandon Campbell, University of Arizona, and Commendations ($500 prize): Bill Elliott, Massachusetts Institute of Technology; David Fannon, Nicole Campbell and Weifang Wang, Rensselaer Polytechnic Institute; Robert Lee, University of Kansas; and Geoffrey Sparks, Matthew Hall and Joshua Bonati, Johns Hopkins University.

A Gallery of Acoustics, sponsored by the Technical Committee on Signal Processing in Acoustics, featured several entries. The winning entry was submitted by Caleb Farny of Boston University.

A short course on Community Noise was presented by Eric Stusnick and Kenneth Plotkin of Wyle Laboratories. The local committee arranged technical tours to the Gibson Guitar plant and The Country Music Hall of Fame.

The plenary session included the annual business meeting of the Society, acknowledgment of the meeting organizers, and announcement and presentation of Society awards and fellowships.

The Medwin Prize in Acoustical Oceanography was presented to Jeffrey A. Nystuen ''for the development of underwater sound measurements to determine rainfall rate and type at sea.'' Earlier in the day Dr. Nystuen gave the Acoustical Oceanography Prize Lecture, as part of the prize program (see Fig. 1). Lori L. Holt was introduced as the recipient of the 2002 Research Grant in Speech Science, sponsored by the American Speech Language Hearing Foundation (see Fig. 2).

The Silver Medal in Psychological and Physiological Acoustics was presented to Brian C. J. Moore of Cambridge University, England, ''for contributions to understanding human auditory perception, especially the perceptual consequences of peripheral frequency analysis in normal and impaired listeners'' (see Fig. 3). The Helmholtz-Rayleigh Interdisciplinary Silver Medal in Underwater Acoustics, Acoustical Oceanography and Signal Processing in Acoustics was presented to Arthur B. Baggeroer of the Massachusetts Institute of Technology ''for applications of model-based signal processing to underwater acoustics and for contributions to Arctic acoustics'' (see Fig. 4). The R. Bruce Lindsay Award was presented to Dani Byrd of the University of Southern California ''for research in motor control and timing in speech production'' (see Fig. 5). The Gold Medal was presented to Richard H. Lyon of RH Lyon Corp, Cambridge, MA, ''for sustained leadership and extensive contributions in the application of statistical concepts to structural acoustics and noise'' (see Fig. 6).

FIG. 1. ASA President Richard Stern (r) presents the Medwin Prize in Acoustical Oceanography to Jeffrey Nystuen (l).

FIG. 2. ASA President Richard Stern (r) congratulates Lori Holt, recipient of the American Speech Language Hearing Foundation's Research Grant in Speech Science.

FIG. 3. ASA President Richard Stern (r) congratulates Brian C. J. Moore, recipient of the Silver Medal in Psychological and Physiological Acoustics.

Election of 15 persons to Fellow grade was announced and fellowship certificates were presented. New fellows are Amy M. Donahue, Hans P. Gottlieb, Jules S. Jaffe, Robert M. Keolian, Oswald Leroy, Raymond Lim, Ronald L. McKay, Zoi-Heleni Michalopoulou, Jens H. Rindel, Robert S. Schlauch, Scott D. Sommerfeldt, Mario A. Svirsky, Floyd E. Toole, David P. Walsh, and Beverly A. Wright (see Fig. 7).

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 D. Wesley Grantham, the Chair of the Meeting, (see Fig. 8), who acknowledged the contributions of the members of his committee including Ralph N. Ohde, Technical Program Chair, Kenneth D. Frampton and Eugene H. Buder, Audio-Visual; Jo-Anne Bachorowski and Benjamin W. Y. Hornsby, Hotel/Facilities; Daniel H. Ashmead, Registration; Corinne M. Darvennes, Technical Tours and Accompanying Persons Program; Mark S. Hedrick, Signs; and Wesley A. Bulla, Publicity. He also extended thanks to the members of the Technical Program Organizing Committee: Ralph N. Ohde, Chair; Ching-Sang Chiu and Bruce Cornuelle, Acoustical Oceanography; Larry Pater, Animal Bioacoustics; Lily Wang and Martha Larson, Architectural Acoustics; Charles Church, Biomedical Ultrasound/Bioresponse to Vibration; Corinne Darvennes, Education in Acoustics; Thomas Howarth, Kim Benjamin and Corinne Darvennes, Engineering Acoustics; Peter Hoekje, Musical Acoustics; Kenneth Cunefare,

FIG. 4. ASA President Richard Stern (l) with Arthur Baggeroer (r), recipient of the Helmholtz-Rayleigh Interdisciplinary Silver Medal.

FIG. 5. ASA President Richard Stern (l) with Dani Byrd (r), recipient of the R. Bruce Lindsay Award.

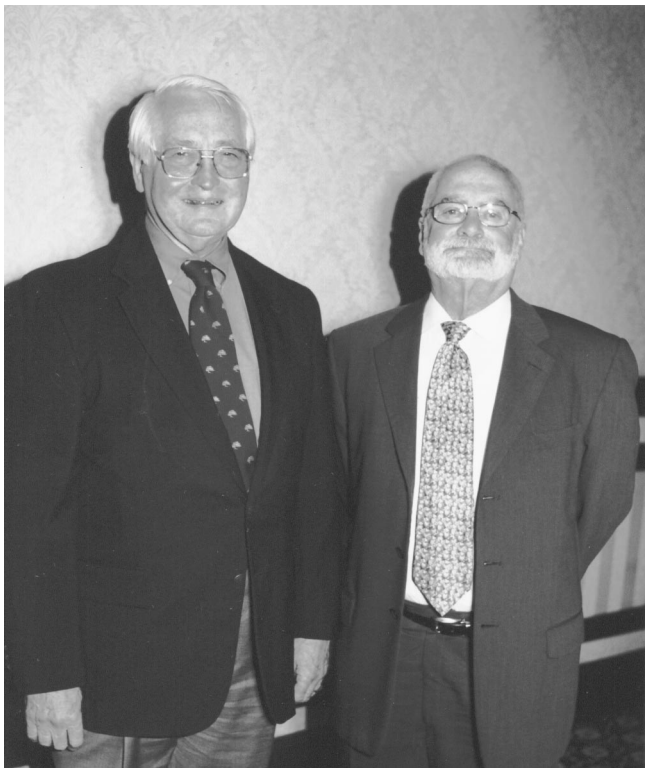


FIG. 6. ASA President Richard Stern (r) with Richard H. Lyon, recipient of the Gold Medal.

FIG. 7. Newly elected fellows of the Society with ASA Vice President William Yost (far right).

FIG. 8. D. Wesley Grantham, Chair of the Nashville meeting.

FIG. 9. William Yost (l) presents the Vice President's gavel to Anthony Atchley (r).

FIG. 10. Ilene Busch-Vishniac (l) presents the President's tuning fork to Richard Stern (r).

Noise; James M. Chambers and Craig Hickey, Physical Acoustics; Ashley Harkrider and Wesley Grantham, Psychological and Physiological Acoustics; Ning Xiang, Signal Processing in Acoustics; Judy R. Dubno and Jo-Anne Bachorowski, Speech Communication; Kenneth Frampton, Structural Acoustics and Vibration; and John Perkins and Peter Gerstoft, Underwater Acoustics.

The Plenary Session also included presentation of the Vice President's gavel by William A. Yost to incoming Vice President Anthony Atchley (see Fig. 9). The session closed with the presentation of the President's tuning fork by Vice President-Elect Ilene Busch-Vishniac to outgoing President Richard Stern (see Fig. 10).

The full technical program and award encomiums can be found in the printed meeting program for readers who wish to obtain further information about the Nashville meeting. We hope that you will consider attending a future meeting of the Society to participate in the many interesting technical events and to meet with colleagues in both technical and social settings. Information on future meetings can be found in the *Journal* and on the ASA Home Page at ⟨http://asa.aip.org⟩.

RICHARD STERN
President 2002–2003

Reports of Technical Committees

Acoustical Oceanography

Fall 2002 Meeting (Cancun). AO sponsored two special sessions. Jeffrey Nystuen organized a session entitled ''Using Ambient Sound in the Ocean,'' and D. Van Holliday, Michael Buckingham, Chester McKinney, and Ian Roebuck organized a session entitled ''David Weston Memorial Session.'' AO in addition co-sponsored a session with UW entitled ''Littoral Environmental Variability and its Acoustic Effects.'' The Best Student Paper Awards in Acoustical Oceanography were awarded to Robert Heitsenrether (University of Delaware), who won first place for his paper ''Influence of fetch limited surface roughness on mid-to-high frequency acoustic propagation in shallow water,'' and Kelly Benoit-Bird (Hawaii Institute of Marine Biology, University of Hawaii), who won second place for her paper ''Measuring acoustic backscattering of deepwater fish in situ using a manned

submersible.'' We are indebted to Timothy Duda (Woods Hole Oceanographic Institution) and Matthew Dzieciuch (Scripps Institution of Oceanography) for representing AO at the Technical Program Organizing Meeting.

Spring 2003 Meeting (Nashville). The 2003 Medwin Prize in Acoustical Oceanography was awarded to Jeffrey Nystuen of the Applied Physics Laboratory, University of Washington, ''for the development of underwater sound measurements to determine rainfall rate and type at sea.'' He gave the Medwin Prize Lecture, entitled ''The Sound of Rainfall at Sea.'' AO sponsored two special sessions. Orest Diachok organized a session entitled ''Bioacoustic Resonance Spectroscopy,'' which included an AO Mini-Tutorial by David T. I. Frances (University of Birmingham, UK) and Kenneth G. Foote (Woods Hole Oceanographic Institution), who spoke on ''Acoustic swimbladder resonance spectroscopy: Fundamentals in scattering theory.'' Peter Gerstoft organized a session entitled ''Geoacoustic Inversion,'' which included an AO Mini-Tutorial by Stan E. Dosso (University of Victoria, Canada), who spoke on ''Probabilistic geoacoustic inversion.'' AO in addition co-sponsored a session with UW entitled ''High Frequency Sediment Acoustics and Associated Sediment Properties and Processes.'' Kelly Benoit-Bird (Hawaii Institute of Marine Biology, University of Hawaii) and Mark Fallat (SACLANT Undersea Research Center, Italy) tied for first place in the competition for the Best Student Paper Award in Acoustical Oceanography. Benoit-Bird's paper was entitled ''Target strength and density structure of Hawaiian mesopelagic boundary community patches,'' and Fallat's paper was entitled ''Geoacoustic characterization of a range-dependent environment using towed array data.'' We are indebted to Ching-Sang Chiu (Naval Postgraduate School) and Bruce D. Cornuelle (Scripps Institution of Oceanography) for representing AO at the Technical Program Organizing Meeting.

Statement of Scope. Following extensive discussions during the AO Technical Committee meetings at the Spring 2002 ASA meeting in Pittsburgh and at the Fall 2002 ASA meeting in Cancun, the AO Technical Committee voted to propose a revised Statement of Scope for the Committee. The concern was that the existing Statement of Scope inappropriately excludes activities that should be included in the field of acoustical oceanography, leaving them without a home in the Society. The revised Statement of Scope, which was approved by the Technical and Executive Councils at the Nashville meeting, is ''The Acoustical Oceanography Technical Committee is responsible for representing and fostering Acoustical Oceanography within the Acoustical Society of America. It is concerned with the development and use of acoustical techniques to measure and understand physical, biological, geological and chemical parameters and processes of the sea.''

AO/UW Special Session Policy. There has been ongoing concern over the extent to which AO and UW sessions sometimes overlap at ASA meetings. AO and UW formed a small joint subcommittee, chaired by James Lynch and Henrik Schmidt, to make recommendations on possible approaches to minimizing this problem. As a result, AO and UW both voted to limit themselves to sponsoring two special sessions per meeting. Occasional exceptions will be allowed for named special sessions honoring an individual, but it is expected that this will be a relatively rare occurrence.

Specialty Subcommittee. At the Cancun meeting AO formed a subcommittee, entitled ''AO Subcommittee on Integrated Acoustics Systems for Ocean Observatories (IASOO).'' The primary purpose of the Committee is to develop the concept of integrated acoustics systems for ocean observatories. The Co-Chairs are Bruce Howe (Applied Physics Laboratory, University of Washington) and James Miller (University of Rhode Island).

Information on these and related matters is available on the Acoustical Oceanography Technical Committee web site, for which we are indebted to Gopu Potty and James Miller. It can be reached through the ASA web page by clicking on ''Committees.'' Please e-mail any material for or comments on the web site directly to potty@oce.uri.edu or miller@oce.uri.edu, with a copy to me at pworcester@ucsd.edu.

I would like to close by thanking all of the volunteers who make the AO Technical Committee a success!

PETER F. WORCESTER
Chair

Architectural Acoustics

Members of TCAA continue to be very dedicated, hard-working, and productive, and our membership continues to grow.

Sixty-one members attended our technical committee meeting in December 2002 in Cancun, Mexico, and 87 members attended our technical committee meeting in April 2003 in Nashville. Highlights of our activities include the ASA meetings, completion of the first major phase of work on classroom acoustics, imminent printing of a book of halls for music performance, support of the second Concert Hall Acoustics Summer Institute (11–14 August 2003 in Saratoga Springs, New York), a student design competition, and continuing work on outreach and education.

Special sessions at the Cancun meeting included design of offices to enhance speech privacy and work performance, acoustical design at hotels, scattering topics in room acoustics, noise isolation, music buildings in Latin America, sound recording and studios in Mexico, ancient acoustics (megaliths and pyramids), halls theaters and cathedrals, speech intelligibility in buildings and metrics for predictions and evaluation, ancient acoustics (vases and theater spaces), ancient acoustics (sounding instruments), ancient acoustics (pre-Colombian sounding instruments), room reverberation modeling, measurements and materials, duct acoustics and vibration, and acoustic comfort in architecture. TCAA also provided a tutorial on architectural acoustics. Session chairs included (note several chaired 2 sessions) Kenneth Roy (2), David Marsh (2), Richard Campbell, Carlos Bidondo, Brandon Tinianov, Christopher Jaffe, Jose Negrete, Murray Hodgson, Frank Brittain, David Lubman, Fernando Elizondo, Peter Mapp, Per Bruel, Sergio Beristain, James Cottingham (2), Roberto Velasquez (2), Angelo Campanella, Fausto Rodriguez (2), Paul Schomer (2), Brigitte Schulte-Fortkamp (2), Jorge Arenas, Jerry Lilly, and Elvira Viverios. The tutorial was chaired by Ralph Muehleisen, with presentations by Neil Shaw, Rick Talaske, and Sylvio Bistafa. Our TPOM representatives were Angelo Campanella and David Marsh.

TCAA members Richard Campbell, Jack Randorff, and Ning Xiang were recognized as new Fellows of ASA. Member Alfred Warnock was awarded the ASA Wallace Clement Sabine Medal. Member Leo Beranek was awarded the John William Strutt, 3rd Baron of Rayleigh Medal by the Mexican Institute of Acoustics.

During the Cancun meeting, Lou Sutherland and David Lubman were presented plaques by Susan Blaeser and Paul Schomer in recognition of outstanding work in connection with the classroom acoustics ANSI standard S12.60.2002, at a reception that was hosted by the National Council of Acoustical Consultants (NCAC).

TCAA member David Lubman was primarily responsible for initiating and organizing the trip on the weekend immediately following the Cancun meeting to Chichen Itza, a compound of astonishing Mayan ruins, including a pyramid and a ball court with intriguing acoustic properties. This overnight adventure involved six buses of ASA members and their families. This fascinating and enjoyable trip seemed an appropriate coda to an energetic week, much dedicated to promoting and enhancing relationships with our friends from the Iberoamerican Federation of Acoustics and the Mexican Institute of Acoustics.

Special sessions at the Nashville meeting included architectural acoustics research and application, multi-channel sound reinforcement systems, student design competition, relationships of synthesis and processing to ''acoustical'' music, a celebration of the life in acoustics of Richard H. Bolt, recording studio acoustics, teaching architectural acoustics to nonacousticians, and architectural and electroacoustic design of interactive museums and halls of fame. Session chairs included William Cavanaugh, Steve Haas, Robert Coffeen, Jeff Szymanski, Alex Case, K. Anthony Hoover, Rendell Torres, Martha Larson, Lily Wang, and Robin Glosemeyer. Special thanks to our TPOM representatives Martha Larson and Lily Wang.

TCAA members Ronald McKay, Floyd Toole, Brigitte Schulte-Fortkamp, and David Walsh were recognized as new Fellows of ASA.

It was announced that ANSI had upheld the classroom acoustics standard S12.60.2002 against all appeals. Plans for outreach and further work were discussed in some detail. Sixteen entries from nine schools competed in the 2003 student design competition, with entries judged by Russ Berger, David Conant, Brian Kubicki, Ed Logsdon, Robert Traub, and Red Wetherill.

Technical initiatives continue to support TCAA's website, student design competition, student receptions, outreach for classroom acoustics, AIA certification for continuing education, and support of the upcoming tutorial on forensics in acoustics (scheduled for the Austin ASA meeting, jointly with NCAC).

Active participation remains strong in TCAA, and our membership continues to grow in both number and enthusiasm.

K. ANTHONY HOOVER
Chair

Animal Bioacoustics

After several years of planning, the First Pan-American/Iberian Meeting on Acoustics in Cancun was a huge success. For members of the Animal Bioacoustics Technical Committee, this meeting opened up new avenues of communication among animal bio-acousticians throughout the world. Animal Bioacoustics sponsored four special sessions in Cancun: Amphibian Bioacoustics in Honor of Robert Capranica organized by Peter Narins; Bioacoustics as a Conservation Tool organized by Renata de Sousa Lima; Coral Reef Acoustics organized by David Mann; and Electrophysiological Investigations of Animals organized by Eduardo Mercado III and jointly sponsored by Psychological and Physiological Acoustics. In addition Arthur Popper and Robert Dooling gave an invited lecture on the History of Animal Bioacoustics. A written version of this lecture will appear in *The Acoustical Society of America—The First 75 Years*, a book being prepared by the History and Archives Committee for distribution at the 75th anniversary meeting in New York in Spring 2004.

At the spring meeting in Nashville, Whit Au organized special sessions on Nature's Orchestra—Acoustics of Singing and Calling Animals, and Animal Bioacoustics jointly sponsored sessions on Bioacoustics Resonant Spectroscopy with Acoustical Oceanography and Underwater Acoustics, and Ecological Psychoacoustics with Psychological and Physiological Acoustics. Two special sessions are planned for the upcoming meeting in Austin: Medical Imaging Techniques to Understand Auditory Processing, and Neurobiology and Communications being organized by Darlene Ketten and George Pollak, respectively.

Congratulations go to winners of this year's student paper awards. At the fall meeting, Jose L. Pena, California Institute of Technology, received the first place award for his paper, ''Genesis of a Space Map in the Owl's Brain,'' and David Blake, University of California at San Francisco, received the second place award for ''Noise Changes Receptive Fields of Auditory Neurons in A1.'' At the spring meeting, Kelly Benoit-Bird, University of Hawaii, received the first place award for presenting ''Behavior of Dusky Dolphins Foraging on the Deep-Scattering Layer in Kaikoura Canyon, New Zealand,'' and Marla Holt, University of California at Santa Cruz, received the second place award for ''Sound Localization of Aerial Broadband Noise in Pinnipeds.''

The First International Conference on Acoustic Communication by Animals will be held 27–30 July 2003, at the University of Maryland—College Park. Arthur Popper, Whitlow Au and Christopher Clark are sharing the responsibilities of organizing and chairing this meeting, which will have over 120 paper presentations by scientists from all parts of the world. This conference was conceived and planned with ASA support as an Animal Bioacoustics technical initiative by Whit Au and Charles Schmid. Many thanks go to all who have contributed to its success.

Members and friends of the Animal Bioacoustics Technical Committee met in December at the Cancun meeting and in April at the Nashville meeting with approximately 35 attending in Cancun and 15 in Nashville. Animal Bioacoustics continues to be served by many excellent volunteers. David Mann attended the Technical Program Organizing Meeting (TPOM) for Cancun, and Larry Pater represented Animal Bioacoustics at the Nashville TPOM. Charles Greene is the Animal Bioacoustics representative to the Medals and Awards Committee and Darlene Ketten serves as our representative on the Membership Committee. Ann Bowles is the Animal Bioacoustics representative to ASA Committee on Standards (ASACOS). Kelly Benoit-Bird is our new representative to the ASA Student Council. She replaces Micheal Dent who was instrumental in getting the Student Council started. Last, but not least, special thanks go to all the session organizers and speakers who contributed to the success of the Animal Bioacoustics program during the past year. Your help is greatly appreciated and our program would not exist without your tremendous effort.

Andrea Simmons was elected as Chair of the Animal Bioacoustics Technical Committee for 2003–2006. It will be an exciting time that includes the 75th anniversary meeting in New York and preparations for the next joint meeting with the Japanese Acoustical Society in Fall 2006. We wish her the very best and look forward to new growth under her leadership.

MARDI C. HASTINGS
Chair 2000–2003

Biomedical Ultrasound/Bioresponse to Vibration

The Fall meeting in Cancun was unusual for the Biomedical Ultrasound/Bioresponse to Vibration (BU/BV) Technical Committee in that it did not focus on a Topical Meeting. The special status of the meeting, with its interactions with acousticians from Latin America, was motivation for organizing a meeting with classical sessions. The BU/BV TC was very active at the meeting with Tuesday afternoon being the only period without a BB session. On Tuesday morning Donald Baker gave a history lecture on Biomedical Ultrasound. There were seven BU/BV sessions: four special sessions, ''Acoustic Microscopy'' organized by Joie P. Jones (co-sponsored by SP), ''Lithotripsy I and II'' organized by Robin Cleveland and Achim Loske, and ''Ultrasound Mediated Drug Delivery and Gene Transfection'' organized by Junru Wu, and three regular sessions. The attendance peaked above 40 in all the sessions. The representative to the Technical Program Organizing Meeting was Jeff Ketterling.

The Technical Committee meeting had 35 people in attendance. Wesley Nyborg was presented with a plaque by Susan Blaeser in honor of his work in leading the technical panel that assisted with the development of the cavitation detection standard. The winners of the student paper competition were Parag Chitnis and Javier van Cauwelaert, both from Boston University. At the Cancun Meeting BU/BV student representative Tyrone Porter (University of Washington) was selected as the F. V. Hunt Postdoctoral Fellow. He will work with Christy Holland at the University of Cincinnati.

At the Nashville meeting BU/BV also had a high profile, with Janet Weisenberger giving a history lecture for Bioresponse to Vibration and Fred Lizzi giving a Hot Topics talk on Ultrasound Therapy. There were five technical sessions: ''Medical diagnostics using sonic frequencies'' organized by Thomas Royston, ''Acoustic Cavitation,'' organized by Carr Everbach (co-sponsored by PA/SP/COS), ''HIFU'' organized by Brian Fowlkes (co-sponsored by PA), ''Image guided therapy'' organized by Ron Roy (co-sponsored by SP), and one regular session chaired by Charlie Church. The TPOM representative was Charlie Church.

There were 42 people in attendance at the Technical Committee meeting at Nashville. Christy Holland gave us a report on the Biomedical Imaging Research Opportunities Workshop (http://www.birow.org/) that was held early in 2003. Christy was the ASA representative to BIROW. The workshop brought together researchers to formulate possible areas which would merit future funding. There were representatives from the National Institutes of Health (NIH) at the meeting who listened to the comments at the meetings. A report on the meeting should be posted on the BIROW web site at the end of this summer. Lawrence Crum was appointed as the new representative to the Medals and Awards Committee. The TC give heartfelt thanks to outgoing representative Wesley Nyborg for his superb service to the Technical Committee on the Medals and Awards Committee. The student paper competition winners were Ajay Anand from University of Washington and Yuan Jing from Boston University.

ROBIN O. CLEVELAND
Chair

Engineering Acoustics

The Engineering Acoustics Technical Committee (EATC) met at each of the two meetings of the Society, in Cancun in the Fall, and in Nashville in the Spring. In Cancun, the EATC sponsored two special sessions and three sessions of contributed papers. The special sessions and their organizers were ''Honoring Per Bruel's Contributions'' organized by Leo Beranek, and ''Metrology Standards and Calibration in Acoustics, Ultrasound and Vibrations'' organized by George S. K. Wong, Jose Salvadore Echeverria and Marc Antonio Nabucode Arauj. In addition, EATC was a co-sponsor of the session ''Underwater Acoustic Measurement Laboratories: New Global Perspectives of Automatic Control and Management,'' in cooperation with the Technical Committee on Underwater Acoustics. The Committee offers many thanks to Elizabeth McLaughlin, who served as the representative to the Cancun TPOM from the Committee. The Committee wishes to recog-

nize James E. West, a longstanding member, for his work as the Meeting Cochair for this first joint meeting with the Iberoamerican Federation of Acoustics and the Mexican Institute of Acoustics, and for his continuing contributions to the Society and the Committee.

The EATC met in Cancun on the evening of Tuesday Dec. 3.

In Nashville, the EATC sponsored one session of contributed papers and co-sponsored one special session. The special session was ''Multichannel Sound Reinforcement Systems,'' a joint session with Architectural Acoustics. The Committee offers many thanks to Tom Howarth and Kim Benjamin, who served as representatives to the Nashville TPOM from the Committee.

The EATC met in Pittsburgh on the evening of Tuesday April 29.

The end of the Nashville meeting marks the end of the term of office for the current Committee Chair. During the spring, a committee chaired by Tom Howarth nominated P. K. Raju and Kim C. Benjamin as candidates for the position of Chair. An election was held, and Kim Benjamin was elected. Kim began his term of office at the end of the Nashville meeting.

The outgoing Chair of EATC would like to thank all of the people who have worked to make the activities of EATC successful for this year, and to invite all interested parties to the EATC meetings to be held at future meetings of the Society.

STEPHEN C. THOMPSON
Chair 2000–2003

Musical Acoustics

2002–2003 was another eventful year for the Technical Committee on Musical Acoustics (TCMU). There were some excellent special technical sessions in musical acoustics at the Cancun meeting in December 2002. These included ''Ancient Acoustics'' organized by David Lubman and Roberto Velazquez (a joint session with Architectural Acoustics), ''Guitar Acoustics'' organized by Tom Rossing, ''Wind Instrument Measurement Techniques'' organized by Peter Hoekje, ''Musical Instruments of the South American Dance Tradition'' organized by Paul Wheeler, and ''Analysis, Synthesis, Perception, and Classification of Musical Sounds'' organized by Jim Beauchamp. In addition to these, the technical committee was asked by the meeting organizers to have a special session (cosponsored by Education in Acoustics) designed primarily for accompanying persons, both adults and children. The resulting session organized by Jim Cottingham and Leonardo Fuks was ''Hands-on Explorations in Acoustics.'' There was good attendance and participation in this session which has been documented photographically in JASA and ECHOES.

The Nashville meeting included four special sessions in musical acoustics, all cosponsored by other technical committees. Tom Cleveland organized ''Singing Voice Acoustics'' (joint with Speech). Jeff Vipperman of Structural Acoustics and Peter Hoekje co-chaired a joint session, ''Structural Acoustics of Musical Instruments.'' Musical Acoustics was co-sponsor of two sessions organized by colleagues in Architectural Acoustics, ''Relationships of Synthesis and Processing to Acoustic Music'' (Tony Hoover) and ''Recording Studio Acoustics'' (Jeff Szymanski and Alexander Case). In addition, Musical Acoustics provided technical initiative funds to help support the special musical entertainment at the social held in Ryman Auditorium.

In addition to the customary efforts to promote student involvement in musical acoustics through the student paper competition and joint sponsorship of student socials, representatives from Musical Acoustics played an active role in the ASA Student Council. Rachel Romond of Penn State completed her term as Student Council Chair in the spring of 2003, and is retiring from formal activity on the council. Andrew Morrison of Northern Illinois University has been appointed as our new Student Council representative. There were nine entrants in the student paper competition in Cancun and seven entrants in Nashville. The winner of the First Prize Student Paper Award at Cancun was Gregoire Derveaux of Ecole Nationale Supérieure de Techniques Avancées who presented ''Time-domain simulation of a guitar.'' Second Prize went to Tamara Smyth (CCRMA, Stanford University) for ''The syrinx: Nature's hybrid wind instrument.'' First Prize at Nashville was awarded to Andrew Morrison of Northern Illinois University for ''Modal analysis and sound radiation from Caribbean steelpans.'' The Second Prize winner was Jared Grogan of Central Washington University who presented ''An experimental study of changes in the impulse response of a wood plate that is subject to vibrational stimulus.''

The committee has two ongoing technical initiative projects. The first is a project being conducted by Peter Hoekje which involves preparation of a set of demonstrations in musical acoustics to be available on CD-ROM and on the web. The second is a project in which selected papers of current interest in musical acoustics are being translated into English. The editorial board for this project is headed by Tom Rossing and includes Jim Beauchamp, Ian Lindevald, and Jim Cottingham. Tom Rossing was appointed as the representative from Musical Acoustics to the Books+ Committee.

The representatives to the Technical Program Organizing Meetings in 2002–2003 were Jim Cottingham for the Cancun meeting and Peter Hoekje for the Nashville meeting. Representation from the TCMU on ASA committees includes Jim Beauchamp on the Medals and Awards Committee, Uwe Hansen on the Membership Committee, and Ian Lindevald on the Committee on Standards. Ian Lindevald also serves on the Panel on Public Policy. Neville Fletcher and Stephen McAdams serve as JASA Associate Editors for Musical Acoustics. Members appointed or reappointed to serve on the technical committee for terms expiring in 2006 include Anders G. Askenfelt, James W. Beauchamp, Xavier Boutillon, Jonas Braasch, Murray D. Campbell, Rene E. Causse, Antoine J. Chaigne, Nicholas J. Giordano, J. M. Harrison, William M. Hartmann, James P. Pyne, Daniel A. Russell, Punita G. Singh, Paul A. Wheeler, and Shigeru Yoshikawa.

The TCMU web site continues to maintain current information on the technical committee and on musical acoustics in general at http://www.public.coe.edu/~jcotting/tcmu. As part of the effort to keep the web site current, the lists ''Musical Acousticians'' and ''Websites for Teaching Acoustics'' were updated in June 2003.

JAMES P. COTTINGHAM
Chair

Physical Acoustics

The first order of business is to thank the recent chair Robert ''Bonzo'' Keolian for three years of outstanding service to the Physical Acoustics TC (PATC) and the ASA. Robert was very active as chair, initiating the Physical Acoustics Student Awards Competition, as well as the PATC e-mail list. We remind the members that although the e-mail list is mostly used for information dissemination, it is available for job postings.

The Fall meeting was held in Cancun, Mexico. The Cancun venue was exciting, and attracted many participants. The meeting was highly successful due in no small part to our session organizers: Glynn Holt organized the ''Bubbles, drops, and foams'' session; Todd Murray organized the ''Laser ultrasonics'' session; and Richard Weaver organized the ''Coda and other stochastic seismic signals'' session. The TPOM organizers for Cancun, Jeff Ketterling and Jim Chambers, did a magnificent job arranging the sessions. Special thanks also go to the session chair volunteers, who kept the sessions on time. The last of the student award competitions for PA occurred in Cancun. The first place prize went to Nico F. Declercq for his talk entitled ''A theoretical elucidation for the experimentally observed backward displacement of waves reflected from an interface having superimposed periodicity;'' second place was awarded to Andreas Hakansson for his talk ''Acoustical properties of finite sonic crystal slabs.'' Thanks to all who contributed and to all who judged the talks. There are no current plans to continue the student awards competition for PA.

Special sessions for the Spring meeting, held in Nashville, TN, included the ''Sono-(con)fusion'' session, organized by Glynn Holt, Felipe Gaitan, and Tom Matula, and the ''Wind noise in outdoor measurements characterization and reduction'' session, organized by Richard Raspet. The Sono-(con)fusion session was run in a nonstandard way, with a small list of invited speakers followed by panel discussions. It was the first time all the major scientists working on acoustic inertial confinement fusion have been brought together. The discussions and debates were lively and spirited. Look for a synopsis of the event in *Echoes*, Vol. 13, No. 3, Summer 2003.

I want to acknowledge members of Physical Acoustics who volunteer their services on various committees: Jim Sabatier serves on the Medals and Awards committee; Steve Garrett serves on the Membership Committee; and Sameer Madanshetty serves on ASACOS. Associate editors also deserve accolades; their hard work helps keep JASA the premier journal for physical acoustics publications. Current and recent members serving as associate editors include Dale Chimenti, Floyd Dunn, Mark Hamilton, John Harris, Michael Howe, Steve Kargl, Martin Ochmann, Richard Raspet, Ron Roy, Andrew Szeri, Louis Sutherland, and Lonny Thompson.

On a sad note, the passing of Robert Apfel leaves us feeling a little empty. We will all miss him. It was a pleasure to see him accept the Gold medal at the Pittsburgh meeting for ''fundamental contributions to physical acoustics and biomedical ultrasound, and innovative leadership in electronic publishing.'' A celebration of his contributions will take place at the New York meeting.

Finally, it is a pleasure to acknowledge members who are active in leadership roles in the ASA. Richard Stern just finished his tenure as president of the ASA, the current vice-president is Anthony Atchley, the vice-president elect is Mark Hamilton, the editor-in-chief is Allan Pierce, and Ron Roy is a Member of the Executive Council and is Editor of ARLO. Thanks to everyone who helps make the PATC a joy!

THOMAS J. MATULA
Chair

Psychological and Physiological Acoustics

The members of the P&P subsection of ASA continue to contribute to the organizational, scientific, and educational aims of the ASA. This is apparent in the TC report, which is rife with well deserved thanks to its members.

Reflecting the policy of P&P to maintain emphasis on the spring meetings, the 144th Meeting of the Acoustical Society of America, in Cancun, was sparsely attended by the members of P&P. The TC Chair did not attend, and acknowledges the efforts of Marjorie Leek and Donna Neff in assistance regarding the administrative components of the fall meeting. In keeping with P&P policy, there will be no special sessions in Austin.

The spring meeting in Nashville was well attended and busy. There were nine sessions sponsored by P&P, one of which was co-sponsored by Speech Communication and one of which was co-sponsored by Animal Bioacoustics. Of the nine sessions, five were special sessions including two named sessions. Thanks to the organizers of the special sessions: Robert Smith (special session in memory of Evan Relkin), Lynne A. Werner and Prudence Allen (developmental psychoacoustics), Robert D. Frisina (honoring the contributions of Jozef Zwislocki), John G. Neuhoff (ecological acoustics), and Astrid Schmidt-Nielson and Patrice S. Beddor (the funding game). Ashley Harkrider accomplished the challenge of organizing the P&P sessions.

The P&P open meeting, while getting a late start, was full of life. After acknowledging the efforts of the local organizing committee, especially P&P member Wes Grantham, several points of business were discussed. The group voted to maintain a technical session on each day of the upcoming NYC meeting. The other technical committees, however, did not mirror our vote; thus no technical sessions will be scheduled for Wednesday at the NYC meeting. There was additional discussion regarding potential celebratory sessions at the NYC, a meeting at which there will be several events surrounding ASA's 75th Anniversary. The Associate Editors (Bernstein, Leek, Lonsbury-Martin, Shofner, and Viemeister) provided brief reports of their activities. Our representative to the Student Council, Joshua Bernstein, provided an overview of the Student Council meeting, and asked P&P members to consider backing the formation of a new ASA award, one for mentorship.

Six new members of the TC were elected: Marsha Clarkson, David Eddins, Lawerence Feth, Hedwig Gockel, Jennifer Lentz, and Mario Ruggero. Their terms begin after the meeting in spring, 2004 (New York City). Thanks to Barbara Shinn-Cunningham, the election chair, and members of the P&P community who so willingly added their names to the ballot.

P&P joined the society in celebrating the award of the Silver Medal in Psychological and Physiological to Brian C. J. Moore. Brian joins an illustrious group of only six previous medalists, and we look forward to his continuing contributions to psychological acoustics.

Two of P&Ps Associate Editors, Leslie Bernstein and Marjorie Leek, completed their successful terms this year. Their efforts on our behalf are greatly appreciated; the work of our associate editors is important as well as time consuming and difficult. Two new associate editors, Gerald Kidd and Armin Kohlrausch, have been named. Their willingness to contribute is valued by the community.

The P&P Technical Initiatives continued unchanged. The initiatives include travel support for invited speakers, student receptions, and homepage maintenance. Suggestions for uses for these funds, including innovations such as workshops, satellite meetings, etc., are welcome (richards@psych.upenn.edu).

VIRGINIA M. RICHARDS
Chair

Signal Processing Acoustics

Signal and image processing continues to provide a basic set of tools that stretch across the Acoustical Society exemplified by the wide variety of applications employing these methods to extract the desired information from synthesized or measured data. Thus, signal processing is a specialty area that many acousticians apply in their daily technical regimen with great success to achieve project goals.

Another exciting year has passed and a number of activities have been accomplished by the Technical Committee on Signal Processing (SP) in Acoustics. Perhaps the most noteworthy event from our perspective is the election of a new chairman, Dr. Charlie Gaumond of the Naval Research Laboratory, who was elected at the Nashville meeting. Over this year the Technical Committee on Signal Processing (SP) in Acoustics has sponsored many technical sessions jointly with the TCs. The Gallery of Acoustics sponsored and administered annually by SP has become a staple at the ASA meetings. Our Technical Liaisons assure that we are continuously interacting at all levels with the other 12 TCs. The Signal Processing Challenge Problem on the ASA website along with our own new SP website provides a venue for signal processing colleagues to interact with us throughout the ASA. The website (http://www.sao.nrc.ca/ims/asa_sp/StartHere.html) provides the interested member a cross-sectional view of the activities of the TC-SP, its goals and objectives, representatives and its membership. Finally, our Best Young Presenter awards have motivated many of our members to present their work and be recognized.

The Technical Committee on Signal Processing (SP) Acoustics fosters a growing interaction with each of the Technical Committees through the sponsoring of joint sessions and tutorials. This effort was made by the various signal processing related sessions throughout the Cancun and Nashville meetings.

At the 144th ASA meeting in Cancun, Mexico, the SP sponsored or co-sponsored 11 sessions entitled ''Acoustic Microscopy'' (BB/SP)—Chair: Joie P. Jones; ''Feature Extraction and Models for Speech'' (SP)—Co-Chairs: Jose A. Diaz, Shrikanth Narayanan; ''The Coda and Other Stochastic Seismic Signals I'' (PA/SP)—Chair: Richard L. Weaver; ''The Coda and Other Stochastic Seismic Signals II'' (PA/SP)—Chair: Joseph A. Turner; ''Acoustic Automatic Target Recognition I and II'' (SP/UW)—Co-Chairs: Ning Xiang, Armando Andrade; ''Speech and Signal Processing'' (SP)—Co-Chairs: Maria Garcia, John G. Harris; ''Seismic Signal Processing: Detection, Estimation and Inversion Methods'' (SP/PA)—Co-Chairs: Alan W. Meyer, Moyses Zindeluk; ''Underwater Signal Processing'' (SP)—Chair: Charles F. Gaumond; ''Time-Reversal Signal Processing'' (SP)—Chair: Philippe Roux; and ''Digital Signal Processing Techniques'' (SP)—Chair: Alan W. Meyer.

We had our ''Best Young Presenter Award'' in which a cash prize was awarded to Mark Skowronski for his presentation entitled ''Human Factor Cepstral Coefficients.''

At the 145th ASA meeting in Nashville, Tennessee, the SP sponsored or co-sponsored 17 sessions entitled ''Geoacoustic Inversion I and III'' (AO/UW/SP)—Chair: Peter Gerstoft; ''Geoacoustic Inversion II'' (AO/UW/SP)—Chair: N. Ross Chapman; ''Dan Johnson: Retrospective'' (NS/ST/SP/AA)—Co-Chairs: Mary M. Prince, John P. Seiler; ''Speech Processing Techniques'' (SP/SC)—Co-Chairs: Douglas O'Shaughnessy, Deborah M. Grove; ''Subspace Methods for Acoustical Imaging I and II'' (SP/PA)—Co-Chairs: David H. Chambers, Sean K. Lehman; ''Robust Passive Sonar I and II'' (UW/SP/EA)—Co-Chairs: Lisa M. Zurk, Brian H. Tracey; ''Celebration in the Life in Acoustics of Richard H. Bolt'' (AA/NS/PA/SP/UW)—Chair: William J. Cavanaugh; ''Interior Noise in Aircraft and Rocket Fairings'' (SA/SP)—Chair: Robert L. Clark; ''General Topics in Signal Processing in Acoustics'' (SP)—Co-Chairs: Ning Xiang, Ronald A. Wagstaff; ''Statistical and Random Analysis'' (SP)—Chair: Leon H. Sibul; ''Arrays and Beamforming'' (SP)—Chair: Alan W. Meyer; ''High-Intensity Focused Ultrasound (HIFU) and Image-Guided Therapy I'' (BB/SP)—Chair: Ronald A. Roy; ''High-Intensity Focused Ultrasound (HIFU) and Image-Guided Therapy II'' (BB/SP)—Chair: J. Brian Fowlkes; ''Nondestructive Evaluation and Material Characterization'' (EA/SP/PA/SA)—Chair: Sally J. Pardue.

We had our ''Best Young Presenter Award'' in which a cash prize was awarded to Nicole Collison for her presentation entitled ''Localizing a large-dimensional field of sonobuoys.'' The Gallery of Acoustics prize was won by Colin Ferrell.

In a concerted effort to enhance the communications between the TC-SP and the other Technical Committees, Technical Liaisons (TLs) are appointed for each TC. A Technical Liaison is a member of the SP who represents the interests of the group to other Technical Committees (e.g., Underwater Acoustics). The TL would attend the SP meeting as well as his/her particular Technical Committee. He/she would primarily facilitate the CO-SPONSORING of joint sessions with SP (our charter) and the Committee, assuring no technical overlap with other sessions. The following SP members are the current Technical Liaisons:

Acoustical Oceanography	Jean-Pierre Hermand
Animal Bioacoustics	Brian Ferguson
Architectural Acoustics	Deborah Grove/Ning Xiang
Bioresponse to Vibration/Biomedical Ultrasound	Sean Lehman
Engineering Acoustics	Stan Ehrlich
Musical Acoustics	John Impagliazzo
Noise	Dave Evans/Joe Pope
Physical Acoustics	Dave Havelock/Dave Chambers
Psychological and Physiological Acoustics	Open
Speech Communication	Jose Diaz
Structural Acoustics and Vibration	Stuart Bolton/Dave Swanson
Underwater Acoustics	Ed Sullivan/Geoff Edelson

We also would like to recognize Charles Gaumond and Jose Diaz for support of the Student Challenge Problem on acoustical signal processing which is now featured on both the ASA and TC-SP websites. This problem is aimed to motivate (and award a cash prize to) signal processing students to participate in some of our activities.

We would like to recognize our outgoing JASA Associate Editor for Acoustical Signal Processing, John C. Burgess. His efforts often go unnoticed, but the results are clearly visible in a higher quality publication. We also welcome our incoming Associate Editor, Edmund J. Sullivan.

We are participating in Acoustic Research Letters On-line (ARLO) and James V. Candy is the Associate Editor for Acoustical Signal Processing. This is an on-line medium offering the exploitation of color, imaging, sound and various multi-media presentations with a rapid turn-around policy for short articles (6 pages) or letters. It is strictly complementary to the normal full-length JASA articles.

The efforts of David Evans on the Standards Committee are also important to the SP. Our representative on the Medals and Awards Committee and Membership Committee are Stanley Ehrlich and David Havelock (respectively). We would also like to recognize Dave Havelock for recreating, updating, and maintaining our website, another task that requires much effort. Surf to our site and get all of the current tidbits of information and links to other processing websites. Our web address is http://www.sao.nrc.ca/ims/asa_sp/StartHere.html. Our Gallery of Acoustics still thrives through the efforts of Preston Wilson and Sean Lehman. Our Student Council member, Hassan Namarvar, continues to represent the TC-SP in a diligent manner.

We would like to recognize Dr. Arthur B. Baggeroer of the Massachusetts Institute of Technology on his selection as the Helmholtz-Rayleigh Interdisciplinary Silver Medalist and his contributions to underwater acoustic signal processing.

Finally, but most important, we would like to congratulate our new Fellows: Ning Xiang of the University of Mississippi, George Ioup of the University of New Orleans, and Zoi-Heleni Michalopoulou of the New Jersey Institute of Technology.

JAMES C. CANDY
Chair 2000–2003

Speech Communication

The Speech Technical Committee (TC) has been very active this past year. The current members of the Committee are: Mariko Aoki, Shari R. Baum, Freddie Bell-Berti, Catherine T. Best, Suzanne E. Boyce, Ann Bradlow, Dani Byrd, René Carré, Randy L. Diehl, David Dorado, Robert D. Frisina, Bruce Gerratt, John H. L. Hansen, Sarah Hawkins, Mark S. Hedrick, Megan M. Hodge, James Hillenbrand, Lori Holt, Sun-Ah Jun, Hideki Kawahara, Klaus J. Kohler, Jody E. Kreiman, Charissa R. Lansing, Qiguang Lin, Nancy S. McGarr, Diane L. Meador, Christopher A. Moore, Shrikanth S. Narayanan, John J. Ohala, Dwayne Paschall, Ian E. Rogers, Robert Ruiz, Astrid Schmidt-Nielsen, Samuel A. Seddoh, Anu Sharma, Caroline L. Smith, Elaine T. Stathopoulos, Winifred Strange, Mario A. Svirsky, Caroline L. Smith, Emily A. Tobey, Greg S. Turner, Nagalapura Viswanath, and Gary G. Weismer. Anders Lofqvist, Douglas O'Shaughnessy. Randy Diehl, Peter Assmann and Kenneth Grant are serving as our Associate Editors. Other ex-officio members of the Committee are Tessa Bent, Gary Weismer and Maureen L. Stone. Thanks to our paper sorters who arrange the technical programs, Diane Kewley-Port and Bob Port for Cancun, and Judy Dubno and Jo-Anne Bachorowski for Nashville. Special thanks to Speech TC members on the Nashville local organizing committee, Ralph Ohde who was Technical Program Chair, Jo-Anne Bachorowski for Hotel/Facilities and who presented the keynote address at the Fellows Lunch, and Mark Hedrick for Signs.

Medals, Awards, and Fellows

Gary Weismer is the representative for Speech Communication to the Society's Medals and Awards Committee and Maureen Stone is our representative to the Membership Committee which makes decisions regarding Fellows of the Society. At the Nashville meeting we honored Dani Byrd at the University of Southern California as the recipient of the R. Bruce Lindsay Award. We also welcomed Mario Svirsky as a Fellow of the Society.

Research Grants in Speech

There are two research grants administered by the Speech TC, the Raymond H. Stetson Scholarship in Phonetics and Speech Production and the Research Grant in Speech Science. The annual Stetson Scholarship supports a student in graduate study in scientific areas related to the field of phonetics and motor speech production. The past year's winner was Cynthia Clopper, Department of Psychology, Indiana University and the award was $3000. Honorable mention was given to Bronwen Evans, University College of London. James Hillenbrand chaired a subcommittee composed of, Gary Weismer, Kenneth Grant and Jody Kreiman who reviewed the 15 proposals submitted. The Research Grant in Speech Science is a biannual award of $5000 supported by the Dennis Klatt Memorial Fund. This fund resides in the American Speech-Language-Hearing Foundation (ASLHF). Since the Fund's inception the intention is that ASLHF and ASA would jointly sponsor and support this research grant. This past year new procedures for reviewing the proposals were implemented. Diane Kewley-Port, Speech TC chair, selected reviewers and coordinated the award selection with ASLHF. Lori Holt at Carnegie Mellon University awarded the Speech Science grant this spring. ASA members wishing to contribute to the Stetson Scholarship Fund can do so to the Acoustical Society Foundation, Inc. Members who are interested in contributing to the Dennis Klatt Memorial Fund should send their donations to the American Speech-Language-Hearing Foundation.

Student Activities

The Committee sponsors two student activities, a competition with a cash award for best oral or poster presentation and a well-attended evening reception with other technical committees. Over 30 students in Cancun and 15 students in Nashville were considered for the Best Student Paper Award. In Cancun, Paulo Escudero at Utrecht was awarded the first prize of $300 and Paul Dyblala at the University of Texas, Dallas, received the second prize of $200 for the Best Paper Award. The coordinator for the judging in Cancun was Catherine Best at Wesleyan University. In Nashville, On Deshmukh at the University of Maryland was the first prize winner of $300, and Cynthia Clopper of Indiana University received the second prize of $200. Coordinator for judging in Nashville was Suzanne Boyce. Tessa Bent at Northwestern University is our Student Council representative. Current activities include a website, pre-meeting email and student outings at meetings. In Austin their initiative to develop a mentoring award will be realized.

Special Sessions

Many exciting sessions were organized in Cancun, including: Voice Processing and Phonetic Algorithms by Sergio Suarez, Maria Esperanza and

Diane Kewley-Port; Spanish and English in Contact by Ann Bradlow and Winifred Strange; Perception and Production of Hearing Impaired by Mario Svirsky and Jorge Gurlekin; and Feature Extraction and Models for Speech (Joint with Signal Processing, Shrikanth Narayanan and Antionio Diaz. Nashville was a very successful meeting with the following special sessions: Neurally Plausible Models for Speech Perception by Robert Port and Frank Guenther; Man and Machine in Cocktail Parties by Pierre Divenyi and Dan Ellis; Acoustic and Perceptual Characteristics of Special Registers by Sarah Hargus Ferguson; and The Funding Game: Strategies for Different Federal Agencies by Astrid Schmidt-Nielsen and Patrice Beddor.

DIANE KEWLEY-PORT
Chair

Structural Acoustics and Vibration

The Cancun meeting (Fall 2002) offered some first time opportunities for close interaction with colleagues from the Iberoamerican Federation of Acoustics and the Mexican Institute of Acoustics. The meeting seemed to be a popular attraction, perhaps because of the location, but there were also many good papers delivered at the meeting. The structural acoustics and vibration TC sponsored four special sessions at this meeting. ''Reconstruction of acoustic radiation from vibrating structures'' was organized by Sean Wu of Wayne State University, ''Analysis, measurement and control of structural intensity'' was organized by Sabih Hayek of Penn State University, ''Vibration of floors in buildings'' was organized by Eric Ungar of Acentech, and ''Energy methods in vibroacoustics analysis and control'' was organized by Jose Arruda of the Universidad Estadual de Campinas in Brazil and Steve Hambric of Penn State University. Even though the meeting was held in Mexico, we had good student participation in our student paper competition. Tatiana Semenova from Wayne State University was the first place winner for her paper entitled ''On the validity and convergence of the Helmholtz equation least squares solution.'' Cecile Goffaux from Univ. Notre-Dame de la Paix in Belgium was the second place winner for the paper entitled ''Evidence of the existence of phononic band gaps: A practical example of a tunable sound insulation by a periodic device of rods.'' Sean Wu (Wayne State University) and Leon Keer (Northwestern University), both of whom are associated with structural acoustics, were presented as new Fellows of ASA at the Cancun meeting, and we wish to congratulate both of them.

At the Nashville meeting (Spring 2003), the structural acoustics and vibration TC sponsored two special sessions. ''Structural acoustics of musical instruments'' was organized by Jeff Vipperman (University of Pittsburgh) and Peter Hoejke (Baldwin-Wallace College) and was joint with the musical acoustics TC. This session seemed particularly appropriate for the Nashville meeting. The second special session was ''Interior noise in aircraft and rocket fairings,'' and was organized by Robert Clark, Jr. from Duke University. The TC again sponsored a student paper competition, and we had a rather large number of students participate, which we were pleased to see. Anne-Marie Albanese from Georgia Tech University was our first place winner for her paper entitled ''When things go crunch: Gap length effects on a magnetorheological tunable vibration absorber.'' Vesna Damljanovic from the University of Illinois was the second place winner for her paper entitled ''Towards a laser-vibrometry technique for measuring railroad rail stress.'' Scott Sommerfeldt was presented as a new Fellow of ASA at the Nashville meeting.

Lance Locey has been serving as our representative on the ASA Student Council, and has done a good job of helping us be more aware of student issues. We continue to have four associate editors for JASA in structural acoustics: Jerry Ginsberg, Andrew Norris, Richard Weaver, and Earl Williams. Paul Remington is an associate editor for ARLO in structural acoustics. Mauro Pierruci is the representative for structural acoustics on the Medals and Awards Committee, and we appreciate his efforts there. He has been instrumental in getting some of our members nominated for well-deserved awards, and the fruits of his labors will be apparent at the Austin meeting. Courtney Burroughs has served as our representative on the Membership Committee, and Joe Dickey, another member of the structural acoustics and vibration TC has just completed nine years of service as the Chair of the Membership Committee. In addition, Courtney was elected as the new Chair of the structural acoustics and vibration TC, and took over those responsibilities at the end of the Nashville meeting. He will most certainly bring the same dedicated service to this position that has characterized his other involvement in the ASA.

SCOTT D. SOMMERFELDT
Chair 2000–2003

Underwater Acoustics

The main activities of The Underwater Acoustics Technical Committee (UWTC) centered around the Cancun and Nashville meetings.

The UWTC representative to the Cancun Technical Program Organizing Meeting (TPOM) was Michael Collins. Because of a number of special circumstances (including the attractiveness of Cancun!) this meeting held an unusually high number of special sessions that were of interest to the underwater acoustics community. A special session on Geoclutter and Boundary Characterization proved to be so popular that it was split into two sessions: the first cochaired by Charles Holland and John Preston, the second cochaired by Purnima Ratilal and Nicholas Makris. A similar circumstance occurred with a special session on Littoral Environmental Variability and Its Acoustic Effects that was cosponsored by the Acoustical Oceanography TC. This session had to be split into three sessions: the first cochaired by Finn Jensen and James Lynch, the second cochaired by Peter Dahl and James Lynch, and the third cochaired by Timothy Duda and Bruce Pasewark. Another special session on Underwater Acoustic Measurement Laboratories was organized by Carlos Ranz-Guerra. As if this were not enough, the UWTC was also cosponsor of two other special sessions: A Memorial Session honoring the work of David Weston (led by the Acoustical Oceanography Committee) and Acoustic Automatic Target Recognition (led by the Signal Processing Committee).

The UWTC was very pleased to have 19 papers submitted for the student paper competition. The winners were Eric Giddens and Gareth Block. The chairman of the judges committee was Mohsen Badiey.

Peter Gerstoft and John Perkins represented the UWTC at the Nashville TPOM. This meeting featured a Memorial Session in honor of Frederick Tappert, organized by David Palmer. The UWTC was also one of several committees to cosponsor a Memorial Session in honor of Richard Bolt. The UWTC joined the Acoustical Oceanography and Signal Processing Committees to cosponsor a session on Geoacoustic Inversion. There were so many contributed papers for this session that it was split into three sessions: the first chaired by Peter Gerstoft, the second chaired by Ross Chapman, and the third again chaired by Peter Gerstoft. This last session was followed by a lively panel discussion on the status and future of geoacoustic inversion. Another popular special session on Robust Passive Sonar had to be split into two sessions. These sessions were cochaired by Lisa Zurk and Brian Tracey. A special session on High Frequency Sediment Acoustics, cosponsored by the Acoustical Oceanography Committee, was organized by Eric Thorsos. The UWTC was also cosponsor of a special session on Bioacoustic Resonance Spectroscopy organized by Orest Diachok.

As for student activities at Nashville, we were pleased to continue our habit of cosponsoring an evening reception with a group of other Technical Committees. The food was not the best we've had at student receptions, but it was (as usual) a large and lively group. The UWTC also held a student paper competition in Nashville. The winners were Ben Dzikowicz and Curtis Osterhoudt.

Prior to the Cancun meeting, The Underwater Acoustics Technical Committee and the Acoustical Oceanography Technical Committee agreed to form a joint Subcommittee to explore ways to minimize the problems we have been having with an abundance of overlapping sessions that are of interest to both groups. At the Nashville meeting the two Committees agreed to restrict the number of special sessions sponsored by each committee to two. It was further agreed that this rule could be violated for special circumstances—for example, if the two Committees are cosponsoring a named session, such as a memorial session.

Dr. Peter Dahl of the Applied Physics Laboratory of the University of Washington has been elected to serve a three year term as Chairman of the Underwater Acoustics Technical Committee. It should also be noted that Peter Rogers is stepping up to the Chairmanship of the Membership Committee and David Bradley is stepping up to the Chairmanship of the Medals and Awards Committee.

JOHN S. PERKINS
Chair 2000–2003

USA Meetings Calendar

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

2003

5–8 Oct.	IEEE International Ultrasonics Symposium, Honolulu, HI [W. D. O'Brien, Jr., Bioacoustics Research Lab., Univ. of Illinois, Urbana, IL 61801-2991; Fax: 217-244-0105; WWW: www.ieee-uffc.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].
3–7 Aug.	8th International Conference of Music Perception and Cognition, Evanston, IL [School of Music, Northwestern Univ., Evanston, IL 60201; WWW: www.icmpc.org/conferences.html].
15–19 Nov.	148th Meeting of the Acoustical Society of America, San Diego, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

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

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

Members of Technical and Administrative Committees of the Acoustical Society of America

The Technical and Administrative Committees listed below have been appointed by the Executive Council. These appointments, with such changes as may be made by the President from time to time, will be in effect until the Spring meeting of the Society in 2004.

Technical Committees 2003–2004

Acoustical Oceanography

	Term to
Peter F. Worcester, *Chair* to 2004	2004
Mohsen Badiey	2006
Paul E. Chandler	2006
Dezhang Chu	2006
Kerry W. Commander	2006
Christian de Moustier	2006
Stan E. Dosso	2006
Andone C. Lavery	2006
Zoi-Heleni Michalopoulou	2006
Jeffrey A. Nystuen	2006
David R. Palmer	2006
Simon D. Richards	2006
Martin Siderius	2006
Daniela Di Iorio	2005
Matthew A. Dzieciuch	2005
Gary J. Heald	2005
Daniel L. Hutt	2005
Timothy G. Leighton	2005
James H. Miller	2005
Carlos J. Robinson	2005
Daniel Rouseff	2005
Emmanuel K. Skarsoulis	2005
Jerome A. Smith	2005
Gang Yuan	2005
Ching-Sang Chiu	2004
Bruce D. Cornuelle	2004
Orest I. Diachok	2004
Kenneth G. Foote	2004
Andrew M. Forbes	2004
Boris G. Katsnelson	2004
James F. Lynch	2004
Jungyul Na	2004
Stephen A. Reynolds	2004
Mark V. Trevorrow	2004

Ex officio:
Shira L. Brochat, Associate Editor of JASA
Kenneth G. Foote, Associate Editor of JASA
William M. Carey, Associate Editor of JASA
William L. Siegmann, Associate Editor of JASA
Ralph A. Stephen, Associate Editor of JASA
Edmund J. Sullivan, Associate Editor of JASA
Alexandra I. Tolstoy, Associate Editor of JASA
Michael J. Buckingham, member of Medals and Awards Committee
Christopher Feuillade, member of Membership Committee
James H. Miller, member of ASACOS

Animal Bioacoustics

	Term to
Andrea M. Simmons, *Chair* to 2006	2006
Ann E. Bowles	2006
John R. Buck	2006
Christopher W. Clark	2006
William C. Cummings	2006
Charles R. Greene	2006
Mardi C. Hastings	2006
D. Vance Holliday	2006
David A. Mann	2006
Sam H. Ridgway	2006
William C. Burgess	2005
James J. Finneran	2005
Adam S. Frankel	2005
Darlene R. Ketten	2005
Larry L. Pater	2005
Peter L. Tyack	2005
Whitlow W. L. Au	2004
Robert J. Dooling	2004
Robert Hickling	2004
David A. Helweg	2004
Peter M. Narins	2004
James A. Simmons	2004

Ex officio:
Whitlow W. L. Au, Associate Editor of JASA
William M. Carey, Associate Editor of JASA
Kenneth G. Foote, Associate Editor of JASA
William L. Siegmann, Associate Editor of JASA
Ralph A. Stephen, Associate Editor of JASA
Charles R. Greene, member of Medals and Awards Committee
Darlene R. Ketten, member of Membership Committee
Ann E. Bowles, member of ASACOS

Architectural Acoustics

	Term to
K. Anthony Hoover, *Chair* to 2004	2004
Nils-Ake Andersson	2006
Walter Beamer	2006
Leo L. Beranek	2006
Sergio Beristain	2006
Sylvio R. Bistafa	2006
Leslie D. Blomberg	2006
Erica Bowden	2006
David T. Bradley	2006
David Braslau	2006
Todd L. Brooks	2006
Courtney B. Burroughs	2006
Paul T. Calamia	2006
Alexander U. Case	2006
William J. Cavanaugh	2006
Dan Clayton	2006
Elizabeth A. Cohen	2006
Chris L. Davies	2006
Damian Doria	2006
John Erdreich	2006
Robin S. Glosemeyer	2006
Robert D. Hellweg	2006
Murray R. Hodgson	2006
Ian B. Hoffman	2006
Jin Jeon	2006
James A. Johnson	2006
Iraklis E. Lampropoulos	2006
Bertram Y. Kinzey, Jr.	2006
Jerry G. Lilly	2006
Peter A. Mapp	2006
David E. Marsh	2006
Gregory A. Miller	2006
Hideo Miyazaki	2006
Richard A. Moscoso-Bullon	2006
Dori A. Najolia	2006
Matthew A. Nobile	2006
Bruce C. Olson	2006
James E. Phillips	2006
Richard J. Peppin	2006
Scott D. Pfeiffer	2006
Joseph Pope	2006
Roy L. Richards	2006
Carl J. Rosenberg	2006
Kenneth P. Roy	2006
Francisco Ruffa	2006
Melvin L. Saunders	2006
Kevin P. Shepherd	2006
Yasushi Shimizu	2006
Gary W. Siebein	2006
Abigail E. Stefaniw	2006
Christopher A. Storch	2006
Jason E. Summers	2006
Louis C. Sutherland	2006
Jiri Tichy	2006
Nancy S. Timmerman	2006
Brandon D. Tinianov	2006
Rendell R. Torres	2006
Gregory C. Tocci	2006
Lily M. Wang	2006
Alfred C. C. Warnock	2006
George P. Wilson	2006
Christopher N. Blair	2005
John S. Bradley	2005
Christopher N. Brooks	2005
Angelo J. Campanella	2005
Quinsan Ciao	2005
Robert C. Coffeen	2005
Peter D'Antonio	2005
Felicia M. Doggett	2005
Timothy J. Foulkes	2005
Robin S. Glosemeyer	2005
Richard D. Godfrey	2005
Clare M. Hurtgen	2005
J. Christopher Jaffe	2005
Mendel Kleiner	2005
David Lubman	2005
Ralph T. Muelheisen	2005
Michael T. Nixon	2005
Jack E. Randorff	2005
H. Stanley Roller	2005
Brigitte Schulte-Fortkamp	2005
Noral D. Stewart	2005
Ning Xiang	2005
Bennett M. Brooks	2004
Steven M. Brown	2004
Richard H. Campbell	2004
M. David Egan	2004
Jesse J. Ehnert	2004
Donna A. Ellis	2004
Ronald R. Freiheit	2004
Richard M. Guernsey	2004
Mark A. Holden	2004
Dana S. Hougland	2004
Jerald R. Hyde	2004
David W. Kahn	2004
Martha M. Larson	2004
Gary S. Madaras	2004

Charles T. Moritz	2004
Paul B. Ostergaard	2004
Dennis A. Paoletti	2004
Neil A. Shaw	2004
Richard H. Talaske	2004
Ewart A. Wetherill	2004
George E. Winzer	2004
Michael R. Yantis	2004

Ex officio:
Jerry H. Ginsberg, Associate Editor of JASA
Mendel Kleiner, Associate Editor of JASA
Andrew N. Norris, Associate Editor of JASA
Richard L. Weaver, Associate Editor of JASA
Earl G. Williams, Associate Editor of JASA
Ewart A. Wetherill, member of Medals and Awards Committee
Gregory C. Tocci, member of Membership Committee
George E. Winzer, member of ASACOS

Biomedical Ultrasound/Bioresponse to Vibration

	Term to
Robin O. Cleveland, *Chair* to 2005	2005
Shira L. Broschat	2006
Ibrahim M. Hallaj	2006
Joie P. Jones	2006
Inder R. Makin	2006
Michael V. Scanlon	2006
Thomas L. Szabo	2006
Janet M. Weisenberger	2006
Junru Wu	2006
Suk Wang Yoon	2006
John S. Allen	2005
Michael R. Bailey	2005
Charles C. Church	2005
Gregory Clement	2005
Floyd Dunn	2005
E. Carr Everbach	2005
Christy K. Holland	2005
R. Glynn Holt	2005
James A. McAteer	2005
Wesley L. Nyborg	2005
Diane Dalecki	2004
J. Brian Fowlkes	2004
Seyed H. R. Hosseini	2004
Kullervo H. Hynynen	2004
James F. Greenleaf	2004
T. Douglas Mast	2004
Thomas J. Matula	2004
Douglas L. Miller	2004
Ronald A. Roy	2004
Thomas J. Royston	2004
Pei Zhong	2004

Ex officio:
Floyd Dunn, Associate Editor of JASA
William P. Shofner, Associate Editor of JASA
Lawrence A. Crum, member of Medals and Awards Committee
Anthony J. Brammer, member of Membership Committee
Robin O. Cleveland, member of ASACOS

Engineering Acoustics

	Term to
Kim C. Benjamin, *Chair* to 2006	2006
Stanley L. Ehrlich	2006
Gary W. Elko	2006
Robert D. Finch	2006
Guillermo C. Gaunaurd	2006
Thomas R. Howarth	2006
Dehua Huang	2006
Sung Hwan Ko	2006
Theodore J. Mapes	2006
Victor Nedzelnitsky	2006
James M. Powers	2006
P. K. Raju	2006
Yongrae Roh	2006
Ahmet Selamet	2006
Stephen C. Thompson	2006
James E. West	2006
George S. K. Wong	2006
Steven R. Baker	2005
David A. Brown	2005
Stephen C. Butler	2005
Robert D. Corsaro	2005
Stephen E. Forsythe	2005
Brian H. Houston	2005
W. Jack Hughes	2005
Robert M. Koch	2005
L. Dwight Luker	2005
Arnie L. Van Buren	2005
Daniel M. Warren	2005
Joseph F. Zalesak	2005
Mahlon D. Burkhard	2004
James Christoff	2004
Fernando Garcia-Osuna	2004
Charles S. Hayden	2004
Dennis F. Jones	2004
Jan F. Lindberg	2004
Yushieh Ma	2004
Elizabeth A. McLaughlin	2004
Alan Powell	2004
Roger T. Richards	2004
Kenneth D. Rolt	2004
Neil A. Shaw	2004
James Tressler	2004

Ex officio:
Yves H. Berthelot, Associate Editor of JASA
Allan J. Zuckerwar, Associate Editor of JASA
Mahlon D. Burkhard, member of Medals and Awards Committee and member of ASACOS
Thomas R. Howarth, member of Membership Committee

Musical Acoustics

	Term to
James P. Cottingham, *Chair* to 2005	2005
Anders G. Askenfelt	2006
James W. Beauchamp	2006
Xavier Boutillon	2006
Jonas Braasch	2006
Murray D. Campbell	2006
Rene E. Causse	2006
Antoine J. Chaigne	2006
Nicholas J. Giordano	2006
J. M. Harrison	2006
William M. Hartmann	2006
James M. Pyne	2006
Daniel A. Russell	2006
Punita G. Singh	2006
Paul A. Wheeler	2006
Shigeru Yoshikawa	2006
George A. Bissinger	2005
Annabel J. Cohen	2005

Diana Deutsch	2005
Uwe J. Hansen	2005
Roger J. Hansen	2005
Peter L. Hoekje	2005
James H. Irwin	2005
Ian M. Lindevald	2005
Stephen E. McAdams	2005
Gary P. Scavone	2005
Gabriel Weinreich	2005
R. Dean Ayers	2004
Judith C. Brown	2004
Courtney B. Burroughs	2004
Robert D. Collier	2004
George F. Emerson	2004
Bozena Kostek	2004
Barry Larkin	2004
Daniel O. Ludwigsen	2004
Thomas D. Rossing	2004
Julius O. Smith	2004
William J. Strong	2004

Ex officio:
Neville H. Fletcher, Associate Editor of JASA
Stephen E. McAdams, Associate Editor of JASA
Edmund J. Sullivan, Associate Editor of JASA
Allan J. Zuckerwar, Associate Editor of JASA
James W. Beauchamp, member of Medals and Awards Committee
Uwe J. Hansen, member of Membership Committee
Ian M. Lindevald, member of ASACOS

Noise

	Term to
Michael R. Stinson, *Chair* to 2006	2006
Jorge P. Arenas	2006
Sergio Beristain	2006
Leslie D. Blomberg	2006
Bennett M. Brooks	2006
Robert D. Bruce	2006
Angelo J. Campanella	2006
William J. Cavanaugh	2006
Gilles A. Daigle	2006
Patricia Davies	2006
Jesse J. Ehnert	2006
Donna A. Ellis	2006
Tony F. W. Embleton	2006
John Erdreich	2006
David J. Evans	2006
William M. Hartmann	2006
Robert D. Hellweg	2006
David K. Holger	2006
Patrick M. Hurdle	2006
Seong-Woo Kang	2006
William W. Lang	2006
Richard H. Lyon	2006
Alan H. Marsh	2006
Ralph T. Muehleisen	2006
William J. Murphy	2006
Kenneth J. Plotkin	2006
Joseph Pope	2006
Daniel R. Raichel	2006
Brigitte Schulte-Fortkamp	2006
Kevin P. Shepherd	2006
Ben H. Sharp	2006
Scott D. Sommerfeldt	2006
Kerrie G. Standlee	2006
George Wong	2006
James C. Yu	2006
Elliott H. Berger	2005
Ann E. Bowles	2005
Frank H. Brittain	2005
Steven M. Brown	2005
Mahlon D. Burkhard	2005
Sanford Fidell	2005
Lawrence S. Finegold	2005
Samir N. Y. Gerges	2005
Richard D. Godfrey	2005
Lee D. Hager	2005
Murray R. Hodgson	2005
Jerry G. Lilly	2005
David Lubman	2005
George A. Luz	2005
Luc Mongeau	2005
Michael T. Nixon	2005
Matthew A. Nobile	2005
Richard J. Peppin	2005
Robert A. Putnam	2005
Jack E. Randorff	2005
Stephen I. Roth	2005
Kenneth P. Roy	2005
Paul D. Schomer	2005
Brandon D. Tinianov	2005
Nancy S. Timmerman	2005
Gregory C. Tocci	2005
Martin Alexander	2004
John P. Barry	2004
Leo L. Beranek	2004
Arno S. Bommer	2004
Anthony J. Brammer	2004
James O. Buntin	2004
Robert J. Cook	2004
Jim R. Cummins	2004
Kenneth A. Cunefare	2004
Joseph M. Cuschieri	2004
Paul R. Donavan	2004
William D. Gallagher	2004
Gerald C. Lauchle	2004
Miomir Mijic	2004
Thomas Norris	2004
John P. Seiler	2004
Noral D. Stewart	2004
Michael R. Stinson	2004
Alice H. Suter	2004
Louis C. Sutherland	2004
Jiri Tichy	2004

Ex officio:
Kenneth A. Cunefare, Associate Editor of JASA
D. Keith Wilson, Associate Editor of JASA
Allan J. Zuckerwar, Associate Editor of JASA
Nancy S. Timmerman, member of Medals and Awards Committee
John Erdreich, member of Membership Committee
Richard J. Peppin, member of ASACOS

Physical Acoustics

	Term to
Thomas J. Matula, *Chair* to 2005	2005
Henry E. Bass	2006
Yves H. Berthelot	2006
James P. Chambers	2006
David I. Havelock	2006
Murray S. Korman	2006
Philip L. Marston	2006
Philip S. Spoor	2006
Larry A. Wilen	2006

D. Keith Wilson	2006
Evgenia A. Zabolotskaya	2006
Robert T. Beyer	2005
Robin O. Cleveland	2005
Lawrence A. Crum	2005
Kenneth E. Gilbert	2005
Robert A. Hiller	2005
R. Glynn Holt	2005
Bart Lipkens	2005
Ralph T. Muehleisen	2005
Harry Simpson	2005
John S. Stroud	2005
Preston S. Wilson	2005
David T. Blackstock	2004
David A. Brown	2004
John A. Burkhardt	2004
Kerry W. Commander	2004
Bruce C. Denardo	2004
Logan E. Hargrove	2004
D. Kent Lewis	2004
Julian D. Maynard	2004
George Mozurkewich	2004
Lev A. Ostrovsky	2004
Andrea Prosperetti	2004
Neil A. Shaw	2004
Victor W. Sparrow	2004
Richard Stern	2004
Roger M. Waxler	2004

Ex officio:
Yves H. Berthelot, Associate Editor of JASA
Dale E. Chimenti, Associate Editor of JASA
Floyd Dunn, Associate Editor of JASA
Guillermo C. Gaunaurd, Associate Editor of JASA
Mark F. Hamilton, Associate Editor of JASA
John G. Harris, Associate Editor of JASA
Michael S. Howe, Associate Editor of JASA
Martin Ochmann, Associate Editor of JASA
Richard Raspet, Associate Editor of JASA
Andrew J. Szeri, Associate Editor of JASA
Louis C. Sutherland, Associate Editor of JASA
Lonny L. Thompson, Associate Editor of JASA
Sean F. Wu, Associate Editor of JASA
Robert Keolian, member of Medals and Awards Committee
Steven L. Garrett, member of Membership Committee
Sameer I. Madanshetty, member of ASACOS

Psychological and Physiological Acoustics

	Term to
Virigina M. Richards, *Chair* to 2005	2005
Søren Buus	2006
Michelle L. Hicks	2006
Lynne Marshall	2006
Christopher J. Plack	2006
Donal G. Sinex	2006
Magdalena Wojtczak	2006
John F. Culling	2005
Torsten Dau	2005
Armin Kohlrausch	2005
Robert A. Lutfi	2005
Christine R. Mason	2005
Brian C. J. Moore	2005
Bruce G. Berg	2004
Joan M. Besing	2004
Monita Chatterjee	2004
Robert D. Frisina	2004
Andrew J. Oxenham	2004

Ex officio:
Gerald D. Kidd, Jr., Associate Editor of JASA
Armin Kohlrausch, Associate Editor of JASA
Brenda L. Lonsbury-Martin, Associate Editor of JASA
William P. Shofner, Associate Editor of JASA
Neal F. Viemeister, Associate Editor of JASA
Joseph W. Hall, member of Medals and Awards Committee
Ervin R. Hafter, member of Membership Committee
Peggy B. Nelson, member of ASACOS

Signal Processing in Acoustics

	Term to
Charles F. Gaumond, *Chair* to 2006	2006
James V. Candy	2006
William M. Carey	2006
Geoffrey S. Edelson	2006
Stanley L. Ehrlich	2006
Brian Ferguson	2006
Paul J. Gendron	2006
Peter Gerstoft	2006
William M. Hartmann	2006
Kevin D. Heaney	2006
William S. Hodgkiss	2006
Paul D. Hursky	2006
John M. Impagliazzo	2006
Hua Lee	2006
Patrick J. Loughlin	2006
Jens M. Meyer	2006
Hassan Namarvar	2006
Joseph W. Posey	2006
James C. Preisig	2006
Brian D. Rapids	2006
Edmund J. Sullivan	2006
David H. Chambers	2005
David J. Evans	2005
Howard A. Gaberson	2005
Charles F. Gaumond	2005
David I. Havelock	2005
Jean-Pierre Hermand	2005
George E. Ioup	2005
Juliette Ioup	2005
Sean K. Lehman	2005
Zoi-Heleni Michalopoulou	2005
Joseph Pope	2005
Leon H. Sibul	2005
Randall W. Smith	2005
James E. West	2005
Gary R. Wilson	2005
Ning Xiang	2005
Leon Cohen	2004
Jose A. Diaz	2004
Deborah M. Grove	2004
David I. Havelock	2004
Alan W. Meyer	2004
Richard J. Ruhala	2004
Roger W. Schwenke	2004
David C. Swanson	2004
Preston S. Wilson	2004

Ex officio:
Edmund J. Sullivan, Associate Editor of JASA
Stanley L. Ehrlich, member of Medals and Awards Committee
David I. Havelock, member of Membership Committee
David J. Evans, member of ASACOS

Speech Communication

	Term to
Diane Kewley-Port, *Chair* to 2004	2004
Jean E. Andruski	2006
Lynne E. Bernstein	2006
Ocke-Schwen Bohn	2006
Suzanne E. Boyce	2006
Bruce R. Gerratt	2006
Benjamin R. Munson	2006
Peggy B. Nelson	2006
Brad H. Story	2006
Fredericka Bell-Berti	2005
Ann R. Bradlow	2005
Dani M. Byrd	2005
James M. Hillenbrand	2005
Lori L. Holt	2005
Hideki Kawahara	2005
Qiguang Lin	2005
Mario A. Svirsky	2005
Gary G. Weismer	2005
Catherine T. Best	2004
Rene Carre	2004
Robert D. Frisina	2004
Megan M. Hodge	2004
Jody E. Kreiman	2004
Charissa R. Lansing	2004
Shrikanth S. Narayanan	2004
John J. Ohala	2004
Dwayne Paschall	2004
Astrid Schmidt-Nielsen	2004
Samuel A. Seddoh	2004
Winifred Strange	2004
Elaine T. Stathopoulos	2004
Emily A. Tobey	2004

Ex officio:
Peter F. Assmann, Associate Editor of JASA
Randy L. Diehl, Associate Editor of JASA
Kenneth W. Grant, Associate Editor of JASA
Anders Lofqvist, Associate Editor of JASA
Douglas D. O'Shaughnessy, Associate Editor of JASA
Gary G. Weismer, member of Medals and Awards Committee
Maureen L. Stone, member of Membership Committee
Shrikanth S. Narayanan, member of ASACOS

Structural Acoustics and Vibration

	Term to
Courtney B. Burroughs, *Chair* to 2006	2006
Benjamin A. Bard	2006
Alain C. Berry	2006
Kenneth A. Cunefare	2006
Joseph W. Dickey	2006
David Feit	2006
Allison B. Flatau	2006
Guillermo C. Gaunaurd	2006
Karl Grosh	2006
Sabih I. Hayek	2006
Louis A. Herstein	2006
Francis Kirschner	2006
Jean R. Nicolas	2006
Andrew N. Norris	2006
Sally J. Pardue	2006
Allan D. Pierce	2006
Mauro Pierucci	2006
Scott D. Sommerfeldt	2006
Earl G. Williams	2006
Sean F. Wu	2006
John A. Burkhardt	2005
Kenneth D. Frampton	2005
Sunil Mehta	2005
Martin L. Pollack	2005
Paul J. Remington	2005
Jeffrey S. Vipperman	2005
Richard L. Weaver	2005
Kuangcheng Wu	2005
Paul E. Barbone	2004
Jeffrey E. Boisvert	2004
Robert L. Clark	2004
Joseph M. Cuschieri	2004
John A. Fahnline	2004
Jerry H. Ginsberg	2004
Stephen A. Hambric	2004
Peter C. Herdic	2004
Harry Himelblau	2004
Timothy W. Leishman	2004
Jerome E. Manning	2004
Philip L. Marston	2004
James G. McDaniel	2004
Karl M. Reichard	2004
Angie Sarkissian	2004
Victor W. Sparrow	2004
Eric E. Ungar	2004
Nickolas Vlahopoulos	2004

Ex officio:
Jerry H. Ginsberg, Associate Editor of JASA
Andrew N. Norris, Associate Editor of JASA
Richard L. Weaver, Associate Editor of JASA
Earl G. Williams, Associate Editor of JASA
Courtney B. Burroughs, member of Membership Committee
Mauro Pierucci, member of Medals and Awards Committee
Louis A. Herstein, member of ASACOS

Underwater Acoustics

	Term to
Peter H. Dahl, *Chair* to 2006	2006
Ralph N. Baer	2006
Kyle M. Becker	2006
John R. Buck	2006
Nicholas P. Chotiros	2006
Dezhang Chu	2006
David R. Dowling	2006
Steven I. Finette	2006
Roger C. Gauss	2006
Paul C. Hines	2006
Anatoliy N. Ivakin	2006
Finn B. Jensen	2006
Sunny R. Khosla	2006
William A. Kuperman	2006
John J. McCoy	2006
B. Edward McDonald	2006
Kazuhiko Ohta	2006
John S. Perkins	2006
John R. Preston	2006
Purnima Ratilal	2006
Ralph A. Stephen	2006
Eric I. Thorsos	2006
Brian H. Tracey	2006
Alexander G. Voronovich	2006
Lisa M. Zurk	2006
Paul B. Baxley	2005
Shira L. Broschat	2005

Douglas H. Cato	2005
Michael D. Collins	2005
Peter H. Dahl	2005
Grant B. Deane	2005
Gerald L. D'Spain	2005
Peter Gerstoft	2005
Charles W. Holland	2005
Kevin B. Smith	2005
Hee Chun Song	2005
Dajun Tang	2005
Christopher T. Tindle	2005
Alexandra I. Tolstoy	2005
Stephen N. Wolf	2005
Ahmad T. Abawi	2004
Michael G. Brown	2004
Dennis B. Creamer	2004
Christian P. de Moustier	2004
Stanley E. Dosso	2004
George V. Frisk	2004
Stewart A. L. Glegg	2004
Joseph F. Lingevitch	2004
Nicholas C. Makris	2004
Zoi-Heleni Michalopoulou	2004
Marshall H. Orr	2004
Gregory J. Orris	2004
James C. Preisig	2004
Daniel Rouseff	2004
William L. Siegmann	2004

Ex officio:
Shira L. Broschat, Associate Editor of JASA
William M. Carey, Associate Editor of JASA
Kenneth D. Foote, Associate Editor of JASA
William L. Siegmann, Associate Editor of JASA
Ralph A. Stephen, Associate Editor of JASA
Edmund J. Sullivan, Associate Editor of JASA
Alexandra I. Tolstoy, Associate Editor of JASA
John S. Perkins, member of Membership Committee
Eric I. Thorsos, member of Medals and Awards Committee
Arnie L. Van Buren, member of ASACOS

Administrative Committees 2003–2004

Archives and History

	Term to
Henry E. Bass, *Chair* to 2004	2004
Jont B. Allen	2006
Ralph R. Goodman	2006
David I. Havelock	2006
Wesley L. Nyborg	2006
Richard J. Peppin	2006
William J. Strong	2006
Leo L. Beranek	2005
William J. Cavanaugh	2005
Logan E. Hargrove	2005
David T. Blackstock	2004
William W. Lang	2004
Julian D. Maynard	2004
Rosalie M. Uchanski	2004

Books^{+}

	Term to
David L. Bradley, *Chair* to 2005	2005
Jerry H. Ginsberg	2006
Philip L. Marston	2006
Joseph Pope	2006
Robert A. Walkling	2006
Nancy S. McGarr	2005
Jeffrey A. Nystuen	2005
Neil A. Shaw	2005
Emily A. Tobey	2005
Stanley L. Chin-Bing	2004
Gordon E. Martin	2004
Victor W. Sparrow	2004

Ex officio:
Allan D. Pierce, Editor-in-Chief

College of Fellows Steering

	Term to
William J. Cavanaugh, *Chair* to 2004	2004
Joseph W. Dickey	2006
Judy R. Dubno	2006
Jiri Tichy	2006
Janet M. Weisenberger	2006
Stanley L. Ehrlich	2005
E. Carr Everbach	2005
Robert D. Frisina	2005
Daniel R. Raichel	2005
Peter G. Cable	2004
M. David Egan	2004
Uwe J. Hansen	2004
Astrid Schmidt-Nielsen	2004

Education in Acoustics

	Term to
Uwe J. Hansen, *Chair* to 2006	2006
Takayuki Arai	2006
William A. Ahroon	2006
Anthony A. Atchley	2006
Fredericka Bell-Berti	2006
Suzanne E. Boyce	2006
Robert D. Celmer	2006
Annabel J. Cohen	2006
E. Carr Everbach	2006
Thomas B. Gabrielson	2006
Katherine S. Harris	2006
Elizabeth S. Ivey	2006
Joie P. Jones	2006
Maria B. Mody	2006
Amy T. Neel	2006
P. K. Raju	2006
Deborah M. Rekart	2006
Daniel A. Russell	2006
M. Roman Serbyn	2006
Victor W. Sparrow	2006
Emily A. Tobey	2006
Courtney B. Burroughs	2005
Robin O. Cleveland	2005
Kenneth A. Cunefare	2005
D. Michael Daly	2005
Mary Florentine	2005
Logan E. Hargrove	2005
Mardi C. Hastings	2005
Peter L. Hoekje	2005
Darrell R. Jackson	2005
Michel T. T. Jackson	2005
Murray S. Korman	2005
Luc Mongeau	2005
John S. Robertson	2005
James M. Sabatier	2005

Neil A. Shaw	2005
Kevin B. Smith	2005
Ralph A. Stephen	2005
James E. West	2005
Wayne M. Wright	2005
David A. Brown	2004
Robert D. Collier	2004
Corinne M. Darvennes	2004
Margaritis S. Fourakis	2004
Carole E. Gelfer	2004
Douglas R. Jones	2004
Sharon Y. Manuel	2004
Philip L. Marston	2004
T. Muehleisen	2004
Andrew A. Piacsek	2004
Daniel R. Raichel	2004
Sally G. Revoile	2004
Thomas D. Rossing	2004
Ronald A. Roy	2004
Dawn R. Schuette	2004
Scott D. Sommerfeldt	2004
William Thompson, Jr.	2004
Robert A. Walkling	2004
George S. K. Wong	2004

Ethics and Grievances

	Term to
Janet M. Weisenberger, *Chair* to 2005	2005
William J. Cavanaugh	2005
Gerald L. D'Spain	2005
Barbara G. Shinn-Cunningham	2005

Investments

	Term to
Ira Dyer, *Chair* to 2004	2004
Patricia K. Kuhl	2006
Lawrence A. Crum	2005
Richard H. Lyon	2005
David Feit, Treasurer, *ex officio*	

Medals and Awards

		Term to
David L. Bradley, *Chair* to 2004		2004
Joseph W. Hall	Psychological and Physiological Acoustics	2006
Lawrence A. Crum	Biomedical Ultrasound/ Bioresponse to Vibration	2006
Robert Keolian	Physical Acoustics	2006
Nancy S. Timmerman	Noise	2006
Ewart A. Wetherill	Architectural Acoustics	2005
Eric I. Thorsos	Underwater Acoustics	2005
Charles R. Greene	Animal Bioacoustics	2005
Mauro Pierucci	Structural Acoustics and Vibration	2005
James W. Beauchamp	Musical Acoustics	2004
Michael J. Buckingham	Acoustical Oceanography	2004
Mahlon D. Burkhard	Engineering Acoustics	2004
Stanley L. Ehrlich	Signal Processing in Acoustics	2004
Gary G. Weismer	Speech Communication	2004

Meetings

May 2003–November 2003

Dana S. Hougland, *Chair* to 2005
Anthony A. Atchley, Vice President
Sergio Beristain, Fall 2002, Cancun
Courtney B. Burroughs, Spring 2002, Pittsburgh
Damian Doria, Spring 2004, New York
Samir N. Y. Gerges, Fall 2002, Cancun
D. Wesley Grantham, Spring 2003, Nashville
Mark F. Hamilton, Vice President-Elect
Russell Johnson, Spring 2004, New York
William A. Kuperman, Fall 2004, San Diego
Elaine Moran, ASA Office Manager, *ex officio*
Clark Penrod, Fall 2003, Austin
Charles E. Schmid, Executive Director, *ex officio*
James E. West, Fall 2002, Cancun

November 2003–May 2004

Dana S. Hougland, *Chair* to 2005
Anthony A. Atchley, Vice President
Sergio Beristain, Fall 2002, Cancun
Damian Doria, Spring 2004, New York
Samir N. Y. Gerges, Fall 2002, Cancun
D. Wesley Grantham, Spring 2003, Nashville
Mark F. Hamilton, Vice President-Elect
Murray D. Hodgson, Spring 2005, Vancouver
Russell Johnson, Spring 2004, New York
William A. Kuperman, Fall 2004, San Diego
Elaine Moran, ASA Office Manager, *ex officio*
Clark Penrod, Fall 2003, Austin
Charles E. Schmid, Executive Director, *ex officio*
James E. West, Fall 2002, Cancun

Membership

		Term to
Peter H. Rogers, *Chair* to 2006		2006
Uwe J. Hansen	Musical Acoustics	2006
Thomas R. Howarth	Engineering Acoustics	2006
Darleen R. Ketten	Animal Bioacoustics	2006
Maureen L. Stone	Speech Communication	2006
Ervin R. Hafter	Psychological and Physiological Acoustics	2006
Anthony J. Brammer	Biomedical Ultrasound/ Bioresponse to Vibration	2005
Courtney B. Burroughs	Structural Acoustics and Vibration	2005
Burton G. Hurdle	Foreign Members	2005
John Erdreich	Noise	2005
Christopher Feuillade	Acoustical Oceanography	2004
Steven L. Garrett	Physical Acoustics	2004
David I. Havelock	Signal Processing in Acoustics	2004
John S. Perkins	Underwater Acoustics	2004
Gregory C. Tocci	Architectural Acoustics	2004

Public Relations

	Term to
Barbara J. Sotirin, *Chair* to 2006	2006
Paul A. Baxley	2006
Paul Hursky	2006
Elizabeth A. McLaughlin	2006

Jack E. Randorff	2006
Kevin B. Smith	2006
James C. Yu	2006
Gerald D'Spain	2005
Geoffrey F. Edelmann	2005
Ellen S. Livingston	2005
James F. Lynch	2005
Duncan E. McGehee	2005
E. Carr Everbach	2004
Charles F. Gaumond	2004
Christy K. Holland	2004
David Lubman	2004
Andrew A. Piacsek	2004

Allan D. Pierce, Editor-in-Chief, *ex officio*
Elaine Moran, ASA Office Manager, *ex officio*
Charles E. Schmid, Executive Director, *ex officio*
Thomas D. Rossing, Echoes Editor, *ex officio*

Publication Policy

	Term to
Floyd Dunn, *Chair* to 2006	2006
Jont B. Allen	2006
David I. Havelock	2006
James F. Lynch	2005
James H. Miller	2005
Allan J. Zuckerwar	2005
Mark F. Hamilton	2004
Patricia K. Kuhl	2004
Alan Powell	2004
Sigfrid D. Soli	2004

William A. Kuperman, President-Elect, *ex officio*
Allan D. Pierce, Editor-in-Chief, *ex officio*

Regional Chapters

Elizabeth A. McLaughlin, *Chair* to 2005

Elmer L. Hixson	Austin
Angelo J. Campanella	Central Ohio
Dean E. Capone	Central Pennsylvania
Scott D. Pfeiffer (acting)	Chicago
Ernest M. Weiler	Cincinnati
Edwin H. Toothman	Delaware Valley
Gary W. Siebein	Florida
Yves H. Berthelot	Georgia
Timothy J. Foulkes	Greater Boston
Michael J. Anderson	Inland Northwest
Neil A. Shaw	Los Angeles
Hari S. Paul	Madras, India
Sergio Beristain	Mexico City
Marehalli G. Prasad	Metropolitan New York
Roger T. Richards	Narragansett
Noral D. Stewart	North Carolina
Peter F. Assmann	North Texas
James R. Angerer	Northwest
David Lubman	Orange County
Paul A. Baxley	San Diego
David Braslau	Upper Midwest
Charles F. Gaumond	Washington, D. C.
Thomas M. Disch	Wisconsin

Ex officio:
Uwe J. Hansen, Chair, Education in Acoustics
David Feit, Treasurer
Erica E. Bowden, Student Council representative

Rules and Governance

	Term to
Tony F. W. Embleton, *Chair* to 2005	2005
Elaine Moran	2006
Charles E. Schmid	2006
William M. Hartmann	2005
Richard H. Lyon	2005
William J. Cavanaugh	2004
Floyd Dunn	2004

Prizes and Special Fellowships

	Term to
Wayne M. Wright, *Chair* to 2005	2005
Fredericka Bell-Berti	2006
James E. West	2006
Uwe J. Hansen	2005
Anthony A. Atchley	2004
Constantine Trahiotis	2004

Standards

Executive Committee
Paul D. Schomer, Chair (Standards Director)
Robert D. Hellweg, Vice Chair
Susan B. Blaeser, Standards Manager, *ex officio*

S1 Representation
John P. Sieler, Chair S1 and ASA representative on S1
George S. K. Wong, Vice Chair S1 and ASA alternate representative on S1

S2 Representation
Richard J. Peppin, Chair S2
David J. Evans, Vice Chair
Sabih I. Hayek, ASA representative on S2
Bruce E. Douglas, ASA alternate representative on S2

S3 Representation
R. F. Burkard, Chair S3 and ASA representative on S3
Craig A. Champlin, Vice Chair S3 and ASA alternate representative on S3

S12 Representation
Robert D. Hellweg, Chair S12
Robert D. Godfrey, Vice Chair S12
Bennett M. Brooks, ASA representative on S12
Paul D. Schomer, ASA alternate representative on S12

International TAGs (ex officio)
Paul D. Schomer, Chair, U.S. TAG for ISO/TC 43 and ISO/TC 43/SC1
David J. Evans, Chair, U.S. TAG for ISO/TC 108
Victor A. Nedzelnitsky, U.S. Technical Advisor for IEC/TC 29

ASA Technical Committee Representatives
Anthony A. Atchley, Chair of ASA Technical Council, *ex officio*
James H. Miller, Acoustical Oceanography
Ann E. Bowles, Animal Bioacoustics
George E. Winzer, Architectural Acoustics
E. Carr Everbach, Biomedical Ultrasound/Bioresponse to Vibration
Mahlon D. Burkhard, Engineering Acoustics
Ian M. Lindevald, Musical Acoustics
Richard J. Peppin, Noise
Sameer I. Madanshetty, Physical Acoustics
Peggy B. Nelson, Psychological and Physiological Acoustics
David J. Evans, Signal Processing in Acoustics

Shrikanth S. Narayanan, Speech Communication
Louis A. Herstein, Structural Acoustics and Vibration
Joseph M. Zalezak, Underwater Acoustics

ASA Officers
David Feit, Treasurer, *ex officio*
Charles E. Schmid, Executive Director, *ex officio*

Past Chair of ASACOS (ex officio)
Tony F. W. Embleton

Associate Editors for Standards News—JASA (ex officio)
Susan B. Blaeser
George S. K. Wong

Tutorials

	Term to
Yves H. Berthelot and Beverly A. Wright, *Cochairs* to 2005	2005
Gerald D. Kidd	2006
Lily M. Wang	2006
Ann E. Bowles	2005
Gary W. Elko	2005
Fredericka Bell-Berti	2004
Robin O. Cleveland	2004
George V. Frisk	2004

Charles E. Schmid, Executive Director, *ex officio*

Women in Acoustics

	Term to
Peggy B. Nelson, *Chair* to 2006	2006
David T. Bradley	2006
Chi-Feng Chen	2006
Melissa A. Epstein	2006
Katherine Sawicki	2006
Subha Maruvada	2006
Uwe J. Hansen	2005
Mardi C. Hastings	2005
Lily Wang	2005
Lisa Zurk	2005
Lawrence A. Crum	2004
Martha M. Larson	2004
Penelope Menounou	2004
Brigitte Schulte-Fortkamp	2004
Elvira B. Viveiros	2004

Mark F. Hamilton, *ex officio* as Vice President-Elect

JASA Editorial Board

June 2006
S. L. Broschat, Underwater Sound
G. D. Kidd, Jr., Psychological Acoustics
A. Kholrausch, Psychological Acoustics
A. Lofqvist, Speech Production
B. L. Lonsbury-Martin, Physiological Acoustics
J. J. McCoy, Mathematical Acoustics
E. Moran, Acoustical News—USA
T. D. Rossing, Education in Acoustics
J. B. Schneider, Computational Acoustics
V. W. Sparrow, Education in Acoustics
E. J. Sullivan, Acoustic Signal Processing
R. Stern, Electronic Archives and References; Forum
A. I. Tolstoy, Underwater Sound

June 2005
P. F. Assmann, Speech Perception
S. B. Blaeser, Acoustical News—Standards
W. M. Carey, Underwater Sound
K. A. Cunefare, Noise, Its Effects and Control
R. L. Diehl, Speech Perception
N. H. Fletcher, Music and Musical Instruments
K. G. Foote, Underwater Sound
G. C. Gaunaurd, General Linear Acoustics
K. W. Grant, Speech Perception
P. L. Marston, Acoustical Reviews—Books
W. G. Mayer, Acoustical News—International
A. N. Norris, Structural Acoustics and Vibration
D. L. Rice, Acoustical Reviews—Patents
W. P. Shofner, Physiological Acoustics
R. Stephen, Underwater Sound
A. J. Szeri, Ultrasonics, Physical Effects of Sound
D. K. Wilson, Noise: Its Effects and Control
G. S. K. Wong, Acoustical News—Standards
A. J. Zuckerwar, Applied Acoustics; Transduction; Acoustical Measurements

June 2004
W. W. L. Au, Bioacoustics—Animal
Y. H. Berthelot, Ultrasonics, Physical Effects of Sound
D. E. Chimenti, General Linear Acoustics
F. Dunn, Bioacoustics—Biomedical
J. H. Ginsberg, Structural Acoustics and Vibration
M. F. Hamilton, Nonlinear Acoustics
J. G. Harris, Ultrasonics, Physical Effects of Sound
M. S. Howe, Atmospheric Acoustics and Aeroacoustics
M. Kleiner, Architectural Acoustics
S. E. McAdams, Music and Musical Instruments
M. Ochmann, General Linear Acoustics
D. D. O'Shaughnessy, Speech Processing
R. Raspet, Ultrasonics, Physical Effects of Sound
W. L. Siegmann, Underwater Acoustics
L. C. Sutherland, Atmospheric Acoustics and Aeroacoustics
L. L. Thompson, General Linear Acoustics
N. F. Viemiester, Psychological Acoustics
R. L. Weaver, Structural Acoustics and Vibration
E. G. Williams, Structural Acoustics and Vibration

Ad Hoc Committees 2003–2004

International Research and Education

Lawrence A. Crum, *Chair*

Sergio Beristain
Malcolm J. Crocker
Gilles A. Daigle
Samir N. Y. Gerges
Konstantin A. Naugolnykh
Oleg V. Rudenko
Brigitte Schulte-Fortkamp

ARLO Editorial Advisory

Ronald A. Roy, *Chair*

E. Carr Everbach
D. Wesley Grantham
Anders Lofqvist
Philip L. Marston
Andrew N. Norris
Allan D. Pierce
Michael R. Stinson

Online Education

E. Carr Everbach, *Chair*

Paul A. Baxley
David T. Bradley
William M. Hartmann
Elaine Moran
Daniel A. Russell

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).]*

Environmental Urban Noise

Amando Garcia, Editor

WIT Press, Southampton, UK, 2001.
225 pp. Price: $136.00 ISBN: 1-85312-752-3.

Environmental Urban Noise constitutes Volume 8 of *Advances in Ecological Sciences* published by WIT Press. This volume is a welcome compendium of current knowledge of environmental urban noise; it provides the basic concepts necessary to understand the field from the viewpoint of the physics of sound, ratings and descriptors for assessment of noise, measurement methodologies, the effect of urban noise on human health, community response, sources of noises in cities, and the multidisciplinary aspects of noise control.

Cities have become less pleasant to reside or work in due to increasing levels of noise pollution generated by trains, automotive vehicles, sirens, etc. The first chapter, which comprises an introduction to the topic, written by the editor himself, points out that deterioration of the acoustic environment essentially began with the Industrial Revolution and gained momentum with the advent of modern transportation systems and the rise of large urban agglomerations. It is interesting to learn here that Robert Koch, the great German physician, predicted more than 100 years ago "the day will come when mankind will have to fight noise just as vehemently as cholera and pestilence." Chapter 2 by Giovanni Brambilla covers the fundamentals of acoustics, characteristics of wave propagation in air, attenuation, absorption in air and other materials, reflection, sound barriers, definitions of loudness, loudness contours, sound level weightings, statistical centile sound levels (L_{10}, L_{90}, etc.), equivalent sound levels, physical assessment and rating of environmental noise, the effects of meteorological conditions on sound propagation, measurement instruments, spectral contents of noise, and the methods of measuring noise levels. Brambilla also describes various types of noise surveys and characteristics of urban noise.

Perhaps the most valuable section in the text is Chap. 3 by Michael Vallet on the effect of noise on health. The short- and long-term effects of noise are discussed, outlining the categories of adaptation, reversible damages, and irreversible damages. In addition to auditory damages, nonauditory effects include cardiovascular responses, respiratory and digestive system malfunctions, and effects on the endocrine and immune systems (e.g., triggering of the hyperthalamus). Noise has been linked to modifications of the hormone system. Methodologies of research on sleep are discussed thoroughly to a degree rarely encountered elsewhere, accompanied by an overview of recent investigations on the impact of noise on sleeping patterns. Another way of evaluating the noise effect on health entails comparison of consumption of medicaments by people exposed to noise with those situated in nonexposed areas. Studies are ongoing on the number of medical consultations and the use of barbiturates and psychotropic drugs (this in spite of the simple procedures of wearing earplugs and closing windows at night). More recent epidemiological statistics indicate an increase in cardiovascular problems and other ailments, but there is work yet to be done in establishing irrefutable attribution of noise to other health problems, occurrence of minor depressions, absenteeism, and also in developing a single parametric index of the effects of noise on health.

In Chap. 4 Rainer Guski covers the topic of community response to noise. While the emphasis is on the causes of noise and community response to noise in Europe, many of the attributes appear to be applicable to the Western Hemisphere and to more highly developed nations of Asia. More than 40% of the European population resides in "gray areas," where L_{eq} lies in the range of 55 to 65 dBA. Traffic noise and aircraft noise constitute the principal contributors to urban pandemonium. Community responses can occur as effects on public health in the guise of increased blood pressure, ischemic heart diseases, annoyance, difficulties in communication, disturbance of recreational activities, and loss of mental concentration. Only recently children became the subjects of noise studies, ranging from fetal development to performance in schools. Low-altitude military flights are more severe on children, inducing observed increases in physiological and physiological stresses. Noise tolerance within a community may depend on personality traits, sensitivity of respondents to noise, and various hard-to-measure factors.

Chapter 5, "Prediction of Urban Noise" by Miguel Arana, deals with the significance of urban noise prediction for planning purposes, comparison of traditional measurements versus computational modules, basic acoustic characteristics (viz., sound power and directivity patterns), factors affecting outdoor sound propagation (geometric spreading, ground attenuation, atmospheric absorption, barriers, vegetation and reflecting surfaces), characteristics of the principal sources of urban noise, and prediction of urban noise, a fundamental theory for estimating traffic noise, vehicle noise regulations (both European and American) and their evolution.

Chapter 6, the final one by editor Amando Garcia, deals with urban noise control. Vehicle noise attenuation may be achieved by development of quieter vehicles, particularly in order to conform with Directive 70/157/CEE adapted by the European Community I 1970 and amended in 1977; road traffic management through relocation of traffic away from noise-sensitive areas, speed limitations, cutting down traffic volume, restricting the presence of heavy trucks in cities, etc.; road design to include screening, lower highway elevations; special road surfaces; and land use planning with due consideration for insulation of dwellings from traffic noise, installation of noise barriers. Aircraft noise control may include establishment of preferential runways, aircraft noise monitoring, curfews, assignment of flight paths, enactment of landing fees, and land zoning to ensure that only noise compatible neighbors occupy the areas immediate to the airport. Railway operation noise can be attenuated through development of quieter vehicles, minimization of rail/wheel interaction, resilient track support structures, and construction of tunnels and noise barriers. Even train depots and bridge crossings come under scrutiny.

Garbage trucks constitute another major source of city noise. Quieter vehicles and specified hours of collection are the chief means of minimizing annoyance. Construction noises also annoy to a great degree, and they are subject in Europe to noise codes of EEC directive 79/113/EE7 (1979), which specifies the test procedures for construction machines. This directive was followed by a series of other directives requiring EEC approval for sound emission from construction equipment. An ever-growing source of urban noise arises from leisure and recreation activities, e.g., rock concerts, discotheques, "boom cars," automobile racing, etc. At present, many of the rules controlling noise in public places are administrated at the local city level. As one antidote to urban noises, acoustical insulation of buildings necessarily entails walls, windows, and floors. The chapter concludes with an overview of noise regulations in Europe on the basis of noise quality criteria. Among the economic incentives to reduce urban noise level are punitive fines, government support of research on noise attenuating measures, monetary rewards for using quieter equipment, and indemnification of parties affected by noise.

In summary, this text provides a satisfactory overview of all aspects of urban noise, complete with a goodly number of recent references. The book should be of major interest to urban planners, health officials, noise control consultants, epidemiologists, and environmental groups. However, the rather steep cost of this rather slim volume may preclude the wider distribution that it merits.

DANIEL R. RAICHEL
Eilar Associates
2727 Moore Lane
Fort Collins, Colorado 80526-2192

OBITUARIES

Ludwig William Sepmeyer • 1910–2002

Ludwig William (Sep) Sepmeyer, a Fellow of the Society and a member since 1932, died at home in West Los Angeles on 18 July 2002 at the age of 91. He was born in East St. Louis, Missouri on 6 November 1910. At the time of his birth, his father was a professional chef in charge of the dining cars for a transcontinental railroad, and hence the family moved from time to time, generally westward, ending in California.

After completing high school in California, Sepmeyer began his undergraduate education at the Los Angeles campus of the University of California. However, he wanted to study engineering, which was not part of the curriculum at that time at the Los Angeles campus. After completing relevant prerequisite courses, he transferred in 1930 to the University of California at Berkeley, where he majored in Electrical Engineering.

Richard H. Bolt was a classmate of his at Berkeley and sparked his interest in acoustics. He joined the Acoustical Society of America in June 1932.

After graduating from Berkeley, Sepmeyer entered the graduate program in the physics department at the University of California at Los Angeles (UCLA), where he studied architectural acoustics under Vern O. Knudsen. At UCLA, he participated in Robert Gales's research by designing frequency-equalizing circuits.

During the 1940s, Sepmeyer worked for the University of California Division of War Research (UCDWR) in the U.S. Navy's laboratory on Point Loma in San Diego, where Vern Knudsen was Chief Scientist. Among his associates in those years were Robert W. Young and Robert Gales. He was responsible for the recordings of ocean ambient noise, which were used in the development of the Knudsen curves of underwater noise vs sea state. It was during this period that Sepmeyer developed his long interest in the design and construction of electrical (later electronic) filters: low-pass, high-pass, and bandpass of various design implementations. He was elected a Fellow of the Acoustical Society of America in May 1944.

After 1945, he was associated for a time with the U.S. Navy's research office in Pasadena, California, and then with the RAND Corporation in Santa Monica. This was during the early years of the space program, and he performed tasks such as estimating the sound-pressure levels in a missile silo during launch. He later joined the System Development Corporation (SDC) as an electrical engineer. He became interested in the vibrational modes in rooms excited by sounds. Two widely cited papers on this subject were published in the *Journal of the Acoustical Society of America*.

Sepmeyer was very active in the Acoustical Society throughout his life. He was a member of the Committee on Membership (1969–1972), the Technical Committee on Engineering Acoustics (1986–1998), the Technical Committee on Architectural Acoustics (1967–2000 and Chair from 1977 to 1980), and the Committee on Standards (1972–1991). His long-term membership in the Society was recognized with the Silver Certificate (25 years) in 1957 and the Gold Certificate (50 years) in 1982.

He was a very active participant in the Accredited Standards Committees S1 on Acoustics and S12 on Noise. In the 1980s, he led the effort to revise the obsolete 1966 issue of the American National Standard for bandpass filters, subsequently published as ANSI Standard S1.11-1986. He also was part of the Working Group of IEC (International Electrotechnical Commission) Technical Committee 29 that prepared the international standard IEC 61260:1995 for bandpass filters.

In addition to his activities with the Acoustical Society, Sepmeyer was an Initial Member (Board Certified in 1978) of INCE/USA. He was a Fellow of the Audio Engineering Society and an active member of the Los Angeles Electronics Club and of the Hollywood Sapphire Club. For many years, he was a private consultant in acoustics, primarily involved with architectural acoustics and noise control projects.

His research in acoustics and electroacoustics led to innovations for the measurement of acoustic energy density and sound intensity, and for the masking of unwanted sounds in open-plan offices. He made significant contributions to the understanding of large-scale Helmholtz resonators and of variable reverberation in multipurpose auditoria.

His wife of 66 years, Inez Hopkins Sepmeyer, and their daughter Adrienne reside in West Los Angeles.

ALAN H. MARSH

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
HASSAN NAMARVAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, Univ. of Rochester, Rochester, New York 14627*

6,529,444

43.35.Sx FOLDED SAGNAC SENSOR ARRAY

Benjamin A. Vakoc, assignor to The Board of Trustees of the Leland Stanford Junior University
4 March 2003 (Class 367/149); filed 24 July 2001

The device falls in the category of fiber-optic acoustic sensor arrays wherein light is propagated in the arrays and the effects of acoustic signals on the light returning from the arrays are analyzed to determine the characteristics of the acoustic signals. A folded Sagnac fiber-optic acoustic sensor operates in a manner similar to a Sagnac interferometer but uses a common delay path to reduce distributed pickup on down-lead fibers. This array can be used to detect acoustic waves in water. By using a Sagnac interferometer rather than a Mach-Zehnder interferometer, the sensor array is said to have a stable bias point, reduced phase noise, and to allow a broadband signal source to be used rather than requiring a more costly narrow-line laser.—DRR

6,532,323

43.35.Sx ACOUSTO-OTPIC FILTER

Byoung Yoon Kim *et al.*, assignors to Novera Optics, Incorporated
11 March 2003 (Class 385/28); filed in the Republic of Korea 6 June 1997

This tunable acousto-optic filter contains a non-birefringent single mode optical fiber with a longitudinal axis, a core and a cladding surrounding the core. The optical fiber has multiple cladding modes and a single core mode that is guided along the core. An acoustic wave propagation member coupled to the optical fiber conveys an acoustic wave from the proximal end to the distal end and launches a flexural wave in the optical fiber. One or more acoustic wave generators are coupled to the proximal end of the acoustic wave propagation member.—DRR

6,535,665

43.35.Sx ACOUSTO-OPTIC DEVICES UTILIZING LONGITUDINAL ACOUSTIC WAVES

Byoung Yoon Kim *et al.*, assignors to Novera Optics, Incorporated
18 March 2003 (Class 385/28); filed 23 May 2000

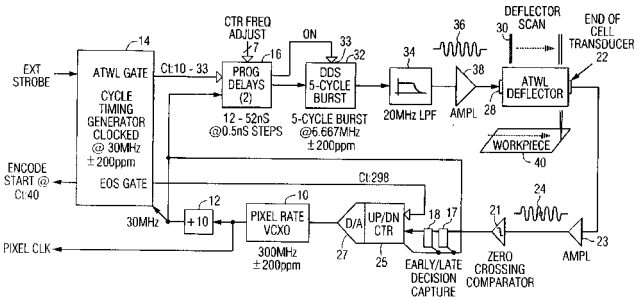
This patent appears to be a very slight variation of United States Patent 6,532,323 reviewed above, awarded to the same people, and assigned to the same company.—DRR

6,538,690

43.35.Sx TIMING CONTROL IN ACOUSTO-OPTIC SCANNER BASED ON ACOUSTIC VELOCITY OF TRAVELING WAVE LENS

Robert M. Montgomery *et al.*, assignors to Harris Corporation
25 March 2003 (Class 348/198); filed 8 March 2000

This is an industrial system for conducting extremely high resolution optical scanning of a workpiece, such as a semiconductor substrate. Critical to successful operation of such a system is the need for extreme precision in the alignment of the light beam and the workpiece. An acoustic traveling wave lens (ATWL) scanner is capable of sustaining such accuracy while scanning at very high speed. The positioning accuracy of the ATWL scanner



derives from the fact that the scan progresses with a traveling acoustic wave in a highly stable material such as fused silica. The present version of this device involves a new method for generating a pixel clock for an ATWL optical scanner wherein the pixel clock is varied in such a manner so as to provide a uniform and constant sampling grid, independently of small acoustic velocity variations in the traveling wave propagation medium.—DRR

6,513,365

43.35.Yb MEASUREMENT OF PHYSICAL CHARACTERISTICS OR PHYSICAL PROPERTIES OF VISCOUS MEDIA BY MEANS OF RAYLEIGH WAVES

Christian Bruetting *et al.*, assignors to Brose Fahrzeugteile GmbH & Company KG, Coburg
4 February 2003 (Class 73/32 A); filed in Germany 13 June 1997

The invention is suited to online measurement of the physical properties of highly viscous liquids. An ultrasonic transmitter sends energy through

a waveguide that contacts the material under test. At least part of the energy passes through a mode converter where it is converted from a Rayleigh wave into a volume sound wave on its way to the receiver. One or more parameters of the Rayleigh wave can then be determined and used to evaluate properties of the test sample. The patent is clearly written and includes 36 pages of illustrations.—GLA

6,537,219

43.35.Yb STATIC FOCUS ULTRASOUND APPARATUS AND METHOD

Mckee D. Poland and Bernard J. Savord, assignors to Koninklijke Philips Electronics N.V.
25 March 2003 (Class 600/447); filed 4 April 2001

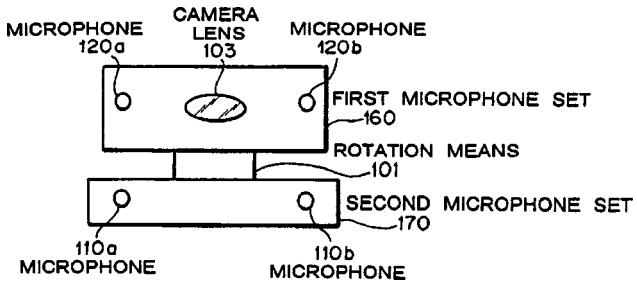
This ultrasonic system uses a transducer assembly that has elements distributed two dimensionally in conjunction with a beam former that, for each beam to be formed, samples the output of each functioning element based on a preset delay value selected for each element that is being used to form the beam in order to generate a c-scan-like set of data.—DRR

6,516,066

43.38.Hz APPARATUS FOR DETECTING DIRECTION OF SOUND SOURCE AND TURNING MICROPHONE TOWARD SOUND SOURCE

Kensuke Hayashi, assignor to NEC Corporation
4 February 2003 (Class 381/92); filed in Japan 11 April 2000

In this video teleconference arrangement, signals received from fixed microphones **110a** and **110b** are processed to compute the angular direction



of the sound source. If the sound is judged to be human speech, then a second pair of microphones **120a** and **120b** are rotated, along with the video camera, to point directly at the talker.—GLA

6,525,993

43.38.Hz SPEAKER DIRECTION DETECTION CIRCUIT AND SPEAKER DIRECTION DETECTION METHOD USED IN THIS CIRCUIT

Yasuhiro Wake and Taisuke Sasada, assignors to NEC Corporation
25 February 2003 (Class 367/127); filed in Japan 23 February 2000

In video teleconference systems an array of two or more microphones can be used to calculate the angular direction of a talker and aim the camera in that direction. However, a simple phase detection circuit becomes easily confused in a noisy environment. The patent describes a signal processing method that involves some fairly sophisticated mathematical algorithms. Search errors are said to be substantially reduced in comparison with prior art.—GLA

6,526,147

43.38.Hz MICROPHONE ARRAY WITH HIGH DIRECTIVITY

Martin Rung, assignor to GN Netcom A/S
25 February 2003 (Class 381/92); filed 12 November 1998

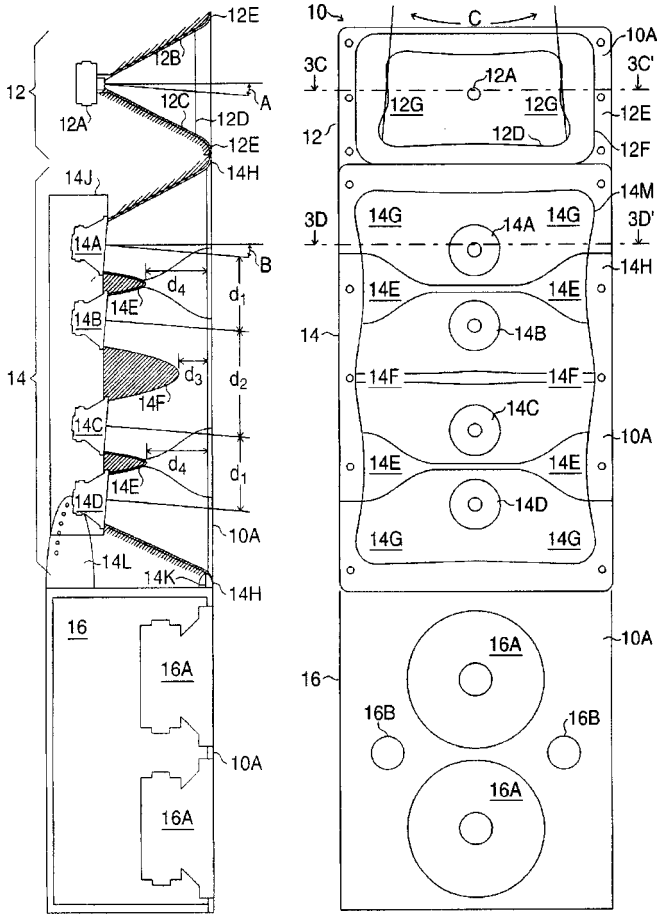
A symmetrical line array of microphones incorporates nonlinear spacing and low-pass filtering. It is known that a loudspeaker array using the same arrangement can achieve uniform directivity over a relatively wide range of frequencies. Most audio engineers would consider the substitution of microphones for loudspeakers a trivial excercise.—GLA

6,513,622

43.38.Ja FULL-RANGE LOUDSPEAKER SYSTEM FOR CINEMA SCREEN

William J. Gelow and Bernard M. Werner, assignors to Harman International Industries, Incorporated
4 February 2003 (Class 181/152); filed 23 August 2000

This patent describes a sophisticated three-way cinema loudspeaker system utilizing special waveguides for the high-frequency and midrange modules. The directivity patterns of these waveguides are designed to compensate for the scattering effect of the perforated screen and to provide uniform sound coverage throughout the seating area of a typical theater. The patent includes design procedures and performance data.—GLA

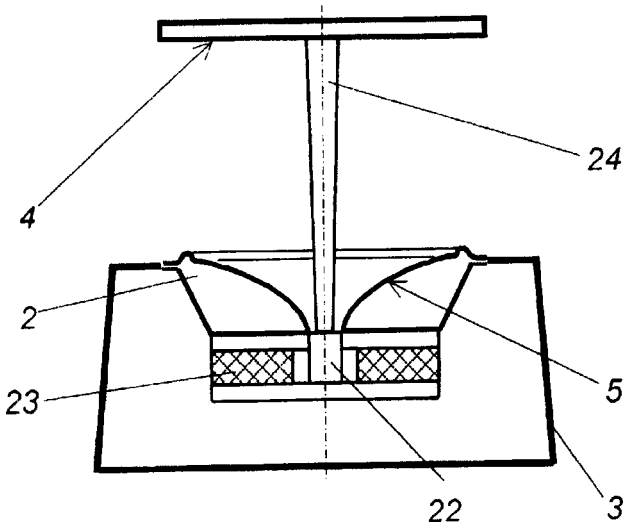


6,516,072

43.38.Ja LOUDSPEAKER

Alexei Vladimirovich Vinogradov and Alexandr Sergeevich Gaidarov, both of Moscow, the Russian Federation
4 February 2003 (Class 381/160); filed in the Russian Federation 11 June 1999

Sound from upward-firing loudspeaker cone **5** bounces off ''axial-symmetric acoustic reflector'' **4** to reduce Doppler distortion (think about it).



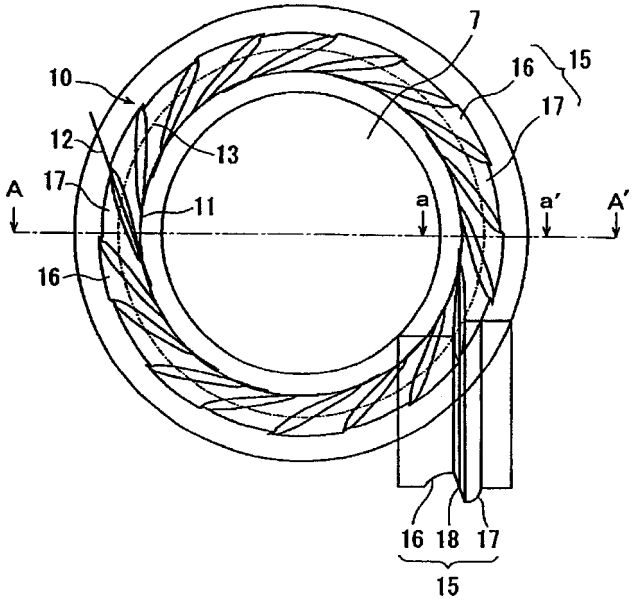
The patent includes a mathematical formula to calculate dimensions in relation to the frequency range of interest.—GLA

6,516,077

43.38.Ja ELECTROACOUSTIC TRANSDUCER

Fumihiko Yamaguchi and Hideo Koreeda, assignors to Foster Electric Company
4 February 2003 (Class 381/424); filed in Japan 1 November 1999

Those familiar with the design of high frequency compression drivers will identify the figure as a variant of the Western Electric tangential diaphragm suspension. In this case, however, the suspension is formed with alternately up-rolled and down-rolled portions. This geometry is said to compensate for uneven driving force and provide ''optimum rigidity of the diaphragm.''—GLA

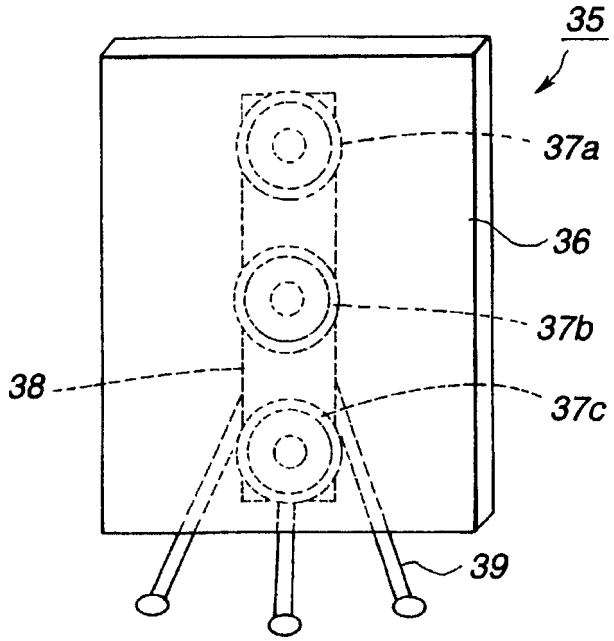


6,519,346

43.38.Ja SPEAKER APPARATUS AND ELECTRONIC APPARATUS HAVING A SPEAKER APPARATUS ENCLOSED THEREIN

Kohei Asada *et al.*, assignors to Sony Corporation
11 February 2003 (Class 381/152); filed 16 September 1999

This patent document runs to more than 50 pages. As described in the single independent claim, however, the basic design consists of only three elements: a free-edge planar diaphragm **36**, two or more driver units **37** that support and energize the diaphragm, and a support member **39**. With suitable signal processing circuitry the device can radiate two wide-range signals in different directions, providing stereophonic reproduction from a single, compact panel.—GLA



6,519,347

43.38.Ja PANEL-FORM LOUNDSPEAKERS

Denis Morecroft and Neil Harris, assignors to New Transducers Limited
11 February 2003 (Class 381/152); filed in the United Kingdom 3 July 1997

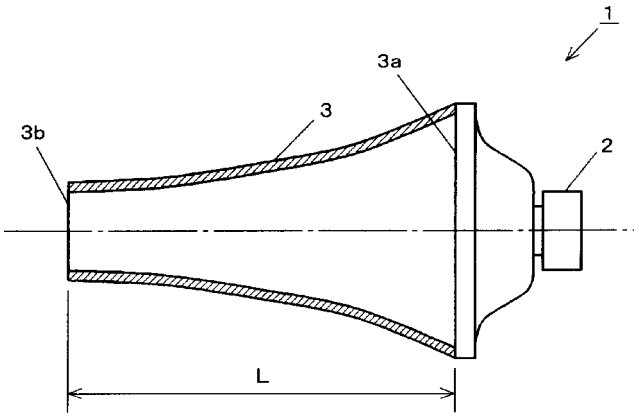
This new invention is a panel-form loudspeaker which has a hole in the panel. Thirty-one claims were granted.—GLA

6,519,348

43.38.Ja SPEAKER APPARATUS AND TELEVISION SET

Kazuhiko Ikeuchi *et al.*, assignors to Matsushita Electric Industrial Company, Limited
11 February 2003 (Class 381/340); filed in Japan 28 May 1998

The shape of a TV picture tube allows a roomy rear chamber for loudspeakers but no extra space at the front for the sound to come out. To



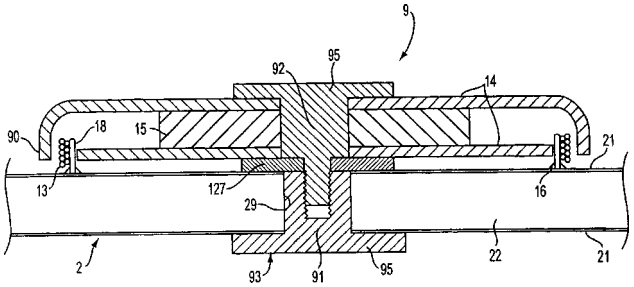
minimize overall cabinet width a waveguide can be used to conduct sound from a round speaker cone to a small vertical slot beside the screen. As would be expected, such reverse-horn waveguides add substantial coloration to the audio signal. However, the inventors are apparently sincere in their belief that ''...it is possible to make an opening of the sound output port smaller in area than an opening of the sound input port in the horn, and at the same time, suppress resonance of sound outputted from the speaker within the horn... .'' No performance measurements are shown.—GLA

6,519,349

43.38.Ja LOUDSPEAKER

Henry Azima *et al.*, assignors to New Transducers Limited
11 February 2003 (Class 381/396); filed in the United Kingdom 2 September 1995

Instead of relying on inertia or an outer frame, this panel-form loudspeaker couples the driving transducer's magnetic assembly **92** to diaphragm **22** at the center of voice coil **13**, somewhat like an oversize cork puller. The forces generated by this arrangement can efficiently launch bending waves into the diaphragm.—GLA

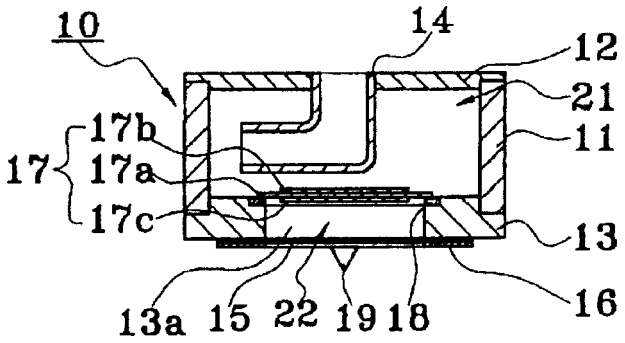


6,522,759

43.38.Ja SPEAKER

Takan Mori, assignor to Murata Manufacturing Company, Limited
18 February 2003 (Class 381/150); filed in Japan 26 December 1997

This small piezoelectric loudspeaker system is designed to provide extended low frequency response. Diaphragm assembly **17** drives tuned rear chamber **21** and front chamber **22**. A ''disc-like film'' **19** covers the front



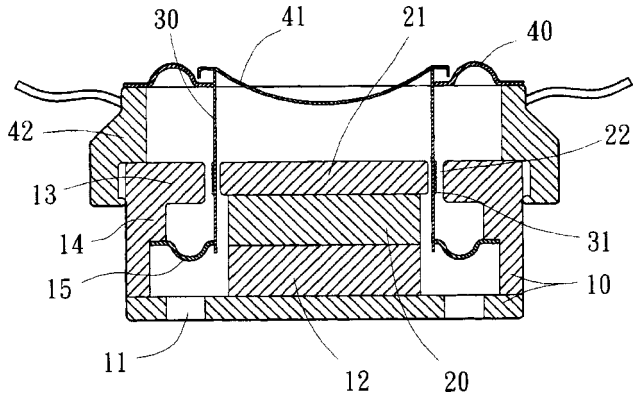
chamber. It appears that the device operates as a familiar push–pull bandpass enclosure in which passive diaphragm **19** takes the place of a second vent.—GLA

6,526,151

43.38.Ja HIGH STABILITY LOUDSPEAKER

Jack Peng, assignor to Meiloon Industrial Company, Limited
25 February 2003 (Class 381/403); filed 29 June 2000

The geometry of this double-spider loudspeaker can easily be visualized from the diagram. One centering device **40** is located above the voice coil and a second centering device **15** is located below the voice coil. Your reviewer can testify that at least one loudspeaker utilizing this basic scheme was built and tested in the late 1960s.—GLA



6,529,611

43.38.Ja MULTIFUNCTION ACOUSTIC DEVICE

Takashi Kobayashi and Akira Yoneyama, assignors to Citizen Electronics Company, Limited
4 March 2003 (Class 381/396); filed in Japan 15 December 2000

This transducer is a combination cell phone vibrator and loudspeaker. The resonant point of the permanent magnet vibrator is set by the compliance of the diaphagm and the spring plate.—MK

6,526,149

43.38.Kb SYSTEM AND METHOD FOR REDUCING NON LINEAR ELECTRICAL DISTORTION IN AN ELECTROACOUSTIC DEVICE

Aleksey S. Khenkin and David E. Blackmer, assignors to Earthworks, Incorporated
25 February 2003 (Class 381/174); filed 28 June 2001

Although other embodiments are described, this patent is concerned mainly with the design of an air gap condenser microphone. The patent teaches that a concave backplate can improve linearity and reduce electrical

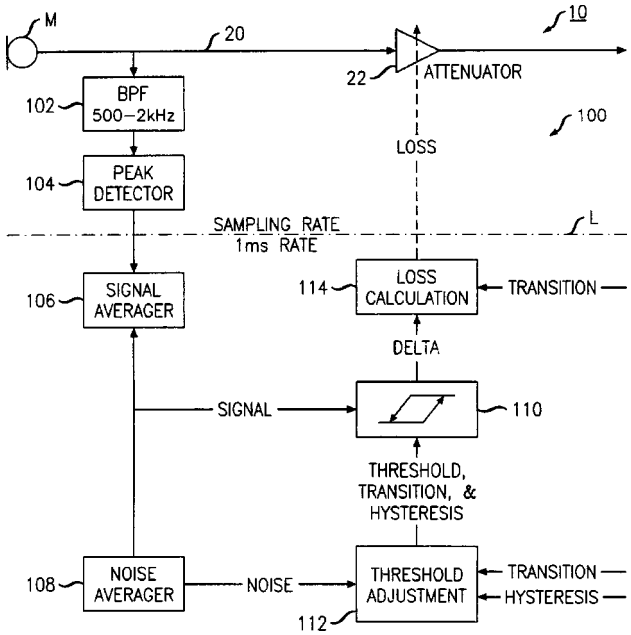
distortion. Alternatively, either the backplate or the diaphragm can be variably polarized to achieve the same result.—GLA

6,516,068

43.38.Lc MICROPHONE EXPANDER

Greg Ciurpita and Scott Pennock, assignors to Lucent Technologies Incorporated
4 February 2003 (Class 381/106); filed 25 August 1998

It has long been known that for telephone speech pickup, the old fashioned carbon button microphone is superior to contemporary dynamic or electret condenser microphones in that it includes built-in compression and



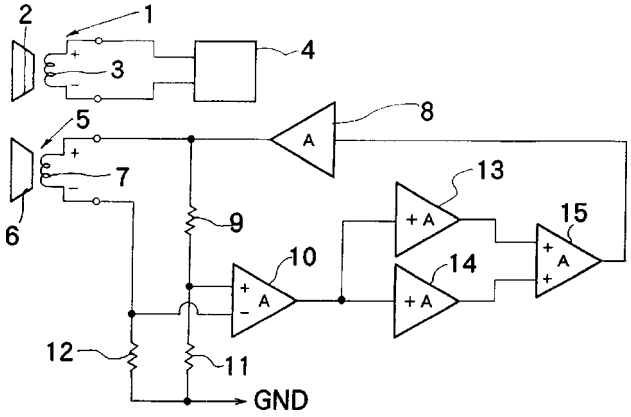
noise reduction. Modern telephone instruments mimic the carbon microphone by adding electronic signal processing. This patent describes an improved digital expander for attenuation of background noise.—GLA

6,522,757

43.38.Lc BASS INTENSIFICATION DEVICE FOR SPEAKER SYSTEM

Shigenori Hiramatsu *et al.*, assignors to Shigenori Hiramatsu; Azden Corporation
18 February 2003 (Class 381/96); filed in Japan 6 June 1997

A closed box houses conventional loudspeaker **1** driven by external power amplifier **4**. It also houses passive radiator **5** which is really a second loudspeaker whose voice coil **7** is connected to the electronic servo circuitry shown. There are no electrical connections between the two devices. Thus, the user's choice of a power amplifier is unhindered.—GLA



6,535,462

43.38.Md DIGITAL SOUND-SIGNAL BROADCASTER

Han-Chih Liu, assignor to Hanpin Electron Company, Limited
18 March 2003 (Class 369/4); filed 15 May 2001

Scratching, or ''turntable music,'' revolves around the use of a second turntable that can be used to introduce new sounds (like scratches) to an existing track. This patent proposes a second turntable that acts as a speed sensor (e.g., via a shaft encoder). This is given to the DSP which can use time compression and expansion algorithms.—MK

6,542,564

43.38.Md METHOD AND APPARATUS FOR COMPENSATING REPRODUCED AUDIO SIGNALS OF AN OPTICAL DISC

Jae Ryong Cho, assignor to LG Electronics Incorporated
1 April 2003 (Class 375/372); filed in the Republic of Korea 4 February 1997

Due to the increased recording density of the DVD format, audio data is recorded at irregular intervals on the disk. So, the playback system must have buffers of sufficient size to store the data during playback. Furthermore, errors could cause buffer underflow and overflow if not detected and corrected.—MK

6,542,613

43.38.Md ELECTRONIC SOUND GENERATOR ATTACHMENT FOR CHILDREN'S SLIDES

Frederick M. Rieber, assignor to Hedstrom Corporation
1 April 2003 (Class 381/124); filed 16 November 1998

The inventor states that ''over time, some children become bored because of the repetitiveness of this play activity'' (sliding down a slide). The solution? Yet another sound chip. This one exhorts the child to continue playing.—MK

6,530,526

43.38.Ne DEVICE FOR TRANSMITTING INFORMATION BY MEANS OF ACOUSTIC SIGNALS

Dietmar Kuschel and Knud Kudnig, assignors to Dietmar Kuschel
11 March 2003 (Class 235/489); filed in Germany 23 December 1997

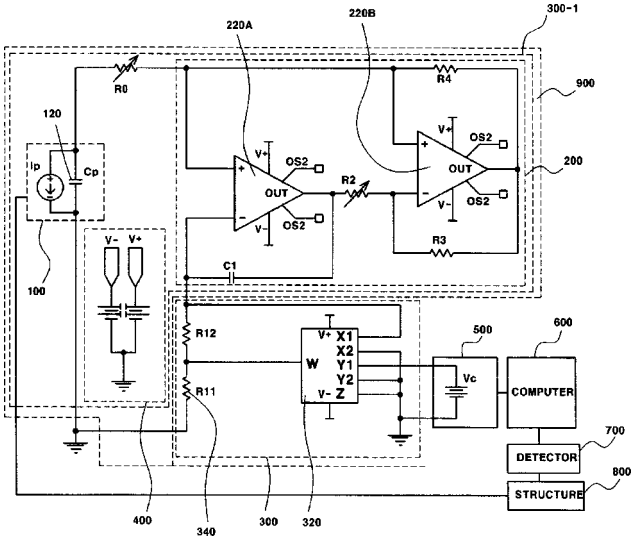
Magnetic stripes are fickle; they wear out due to rubbing, they can easily be demagnetized, etc. So, instead of a magnetic stripe, a column of raised bumps could be used with a microphone.—MK

6,538,401

43.40.Vn ADAPTIVE SHUNT SYSTEM FOR VIBRATION CONTROL OF STRUCTURES AND OPERATING METHOD OF THE SAME

Keun Ho Rew *et al.*, assignors to Korea Advanced Institute of Science and Technology
25 March 2003 (Class 318/114); filed in the Republic of Korea 4 July 2001

The object of the device is to provide an adaptive shunt system for controlling structural vibrations. Vibrations generated in a structure are



detected and processed in real time to generate a command voltage which drives an adaptive shunt device in which the natural frequency varies at every moment according to the voltage command.—DRR

6,533,691

43.50.Jh LOW NOISE SILENT CHAIN

Hiroshi Horie *et al.*, assignors to Tsubakimoto Chain Company
18 March 2003 (Class 474/213); filed in Japan 15 March 2000

Each link plate of this silent chain has an inside surface formed on the inside flank of each link tooth. The inside meshing surface incorporates a convexly arcuate mesh located on the tip side and on the root side of each link tooth. When the chain engages a sprocket, the engagement proceeds more smoothly and continuously from the inside meshing surface to the outside meshing surface, thus resulting in a reduction of noise level.—DRR

6,536,284

43.50.Jh METHOD AND APPARATUS FOR COMPRESSOR CONTROL AND OPERATION VIA DETECTION OF STALL PRECURSORS USING FREQUENCY DEMODULATION OF ACOUSTIC SIGNATURES

Pierino Gianni Bonanni, assignor to General Electric Company
25 March 2003 (Class 73/660); filed 12 June 2001

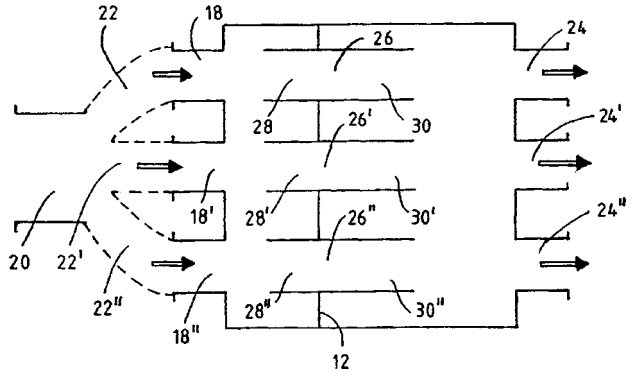
The method relates to noninvasive techniques for monitoring the rotating components of machinery. The compressor pressure ratio of an industrial gas turbine engine is usually set at a predetermined margin away from the surge-stall limit in order to avoid unstable compressor operation. The compressor state is monitored by one or more attached sensors. The sensor results are analyzed to recover stall precursors, which are compared with preset baseline data. A controller initiates corrective actions to prevent compressor surge and stall when the stall precursors deviate from the preset baseline level. The processor system may also incorporate a frequency demodulator and a system for processing the frequency-demodulated signals to extract stall precursor characteristics.—DRR

6,530,452

43.50.Gf REACTIVE SILENCER FOR INDUSTRIAL AIR CHANNELS AND ITS USE

Henrik Pettersson *et al.*, assignors to Metso Paper, Incorporated
11 March 2003 (Class 181/251); filed in Finland 30 September 1998

This is a reactive silencer intended for an industrial air supply or an exhaust air channel. It consists of a sound attenuator chamber divided into two parts by a partition. The partition contains two or more channels or



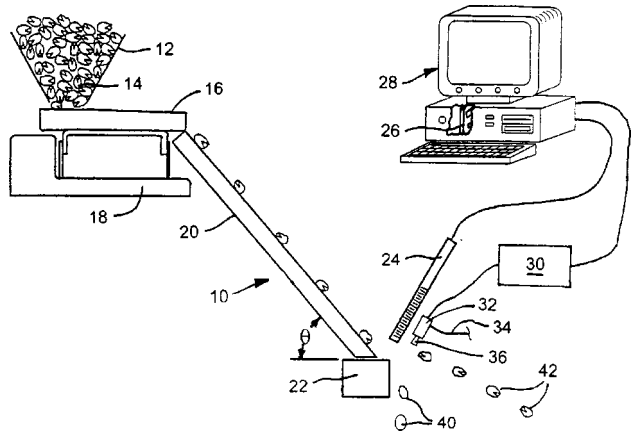
pipes, thus connecting the two parts of the chamber. This is said to result in a silencer of small size and simple construction for noise produced by large or small air channels.—DRR

6,541,725

43.50.Yw ACOUSTICAL APPARATUS AND METHOD FOR SORTING OBJECTS

Thomas C. Pearson, assignor to The United States of America as represented by the Secretary of Agriculture
1 April 2003 (Class 209/590); filed 3 April 2001

This is equipment for automatic sorting of small objects, such as pistachio nuts, using the sound of the falling objects as the basis for sorting. The sorting process begins by bouncing the object off a surface so that the object emits a sound. The sound is then converted into an electrical signal



that is interpreted to indicate the trait of the object. For example, this method can be used to sort pistachio nuts based on whether their shells are open or closed. The electrical signal can be integrated and a signal gradient produced to discriminate among signals indicating different classes of objects.—DRR

6,530,451

43.55.Cs SYSTEM OF ELEMENTS FOR THE DIFFUSION OF SOUND IN ROOMS DELIGATED TO THE REPRODUCTION OF MUSIC AND SPEECH

Guido Noselli, assignor to Outline S.N.C. di Noselli G. & C.
11 March 2003 (Class 181/179); filed in Italy 21 March 2000

The patent describes a system for sound diffusion at low frequencies that includes a large horn enclosure, perhaps of brickwork, and a ''multi-hemicylindrical'' surface that can be adjusted to tune the absorption.—CJR

6,526,704

43.55.Ev PADDED RAISED FLOORING PANELS AND COVERINGS

Raymond A. Berard *et al.*, assignors to Interface, Incorporated
4 March 2003 (Class 52/177); filed 22 December 2000

These raised access flooring panels incorporate a thick resilient pad bonded directly to the panel. The pad absorbs impact noise. A finished flooring such as carpet can be added on top of the pad and easily removed without having to change the entire panel.—CJR

6,530,453

43.55.Ev SOUND ABSORBING STRUCTURE

Hiroaki Miura *et al.*, assignors to Nissan Motor Company, Limited
11 March 2003 (Class 181/292); filed in Japan 11 November 1999

This sound absorbing structure comes from a single resin mold. It has multiple openings to interior voids that provide resonance absorption as well as a surface membrane for sound absorption. The structure could be used for a vehicular body part.—CJR

6,531,541

43.55.Ev COATING COMPOSITIONS, COATED SUBSTRATES AND METHODS FOR INHIBITING SOUND TRANSMISSION THROUGH A SUBSTRATE

Umesh C. Desai *et al.*, assignors to PPG Industries Ohio, Incorporated
11 March 2003 (Class 524/832); filed 19 May 2000

The invention describes coating compositions that can be sprayed. They include an aqueous dispersion of polymeric microparticles, which are useful for sound deadening applications and vibration damping, and are beneficial in the automotive industry.—CJR

6,541,105

43.55.Ev ACOUSTICAL OPEN-CELL POLYLEFINS AND PROCESS FOR MAKING

Chung P. Park, assignor to Dow Global Technologies Incorporated
1 April 2003 (Class 428/304.4); filed 24 January 2002

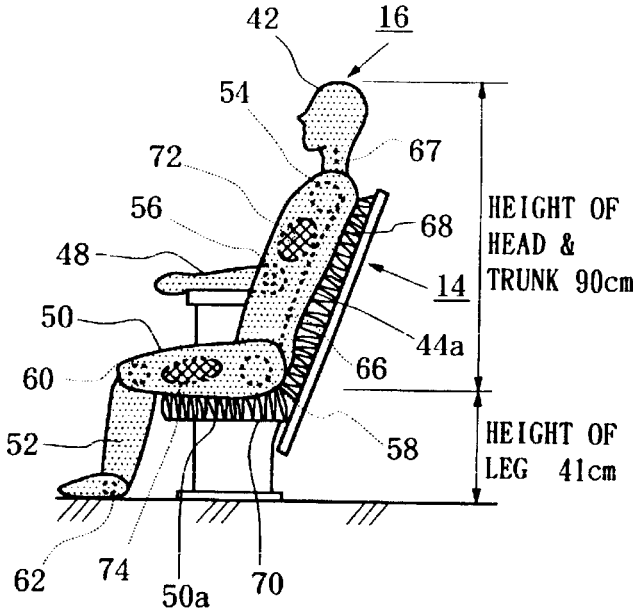
The invention provides an expanded open-cell foam polymer composition suitable for sound absorption in automotive and appliance applications.—CJR

6,463,817

43.55.Gx DUMMY SOUND ABSORBER REPRESENTING A SEATED HUMAN FOR MEASURING SOUND ABSORPTION POWER OF A THEATER CHAIR

Takeshi Sakai and Kawakami Fukushi, assignors to Yamaha Corporation
15 October 2002 (Class 73/865.8); filed in Japan 1 October 1992

This patent describes a human-shaped sound absorber for use in testing music hall or auditorium acoustics. A basic styrofoam body shape may be



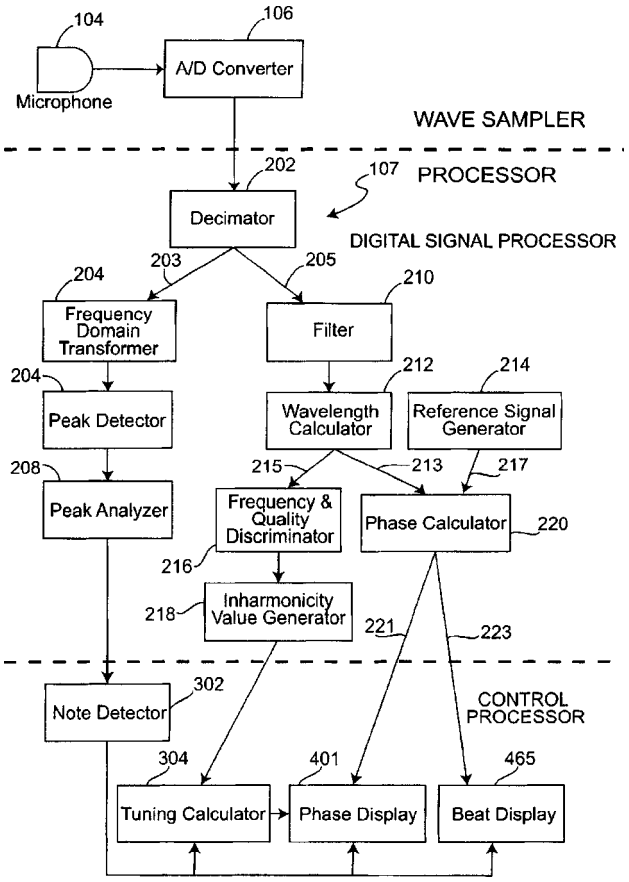
covered with an appropriate surface layer. Weights **72**, **74** may be added to compress the seat in a realistic manner. The patent provides considerable information on impulse-type analysis methods for concert hall acoustics.—DLR

6,529,843

43.58.Hp BEAT RATE TUNING SYSTEM AND METHODS OF USING SAME

David J. Carpenter, Chicago, Illinois
4 March 2003 (Class 702/75); filed 12 April 2001

It is well known that piano strings exhibit inharmonicity. This invention is designed to detect the inharmonic partials of each piano note and



offer corrections by creating a beat display. The patent is well written and appears viable.—MK

6,532,005

43.58.Ta AUDIO POSITIONING MECHANISM FOR A DISPLAY

Lowell Campbell, assignor to DENSO Corporation
11 March 2003 (Class 345/173); filed 17 June 1999

This audio user interface is particularly simple: volume in Y, frequency in X. The lack of prior patent citations suggests that the inventor is unaware of the audio display community.—MK

6,540,613

43.58.Ta VIDEO GAME APPARATUS, BACKGROUND SOUND OUTPUT SETTING METHOD IN VIDEO GAME, AND COMPUTER-READABLE RECORDING MEDIUM STORING BACKGROUND SOUND OUTPUT SETTING PROGRAM

Satoru Okubo and Shigeru Araki, assignors to Konami Corporation
1 April 2003 (Class 463/35); filed in Japan 13 March 2000

Simply put, background sound should depend on the circumstances of the computer game. Accordingly, depending on the position of the viewpoint, the sounds should change in loudness and character. Only then will the listener have ''a game full of feeling.'' Surely this is obvious to those ''skilled in the art.''—MK

6,538,570

43.60.Lq GLASS-BREAK DETECTOR AND METHOD OF ALARM DISCRIMINATION

Richard A. Smith, assignor to Honeywell International
25 March 2003 (Class 340/550); filed 22 June 2000

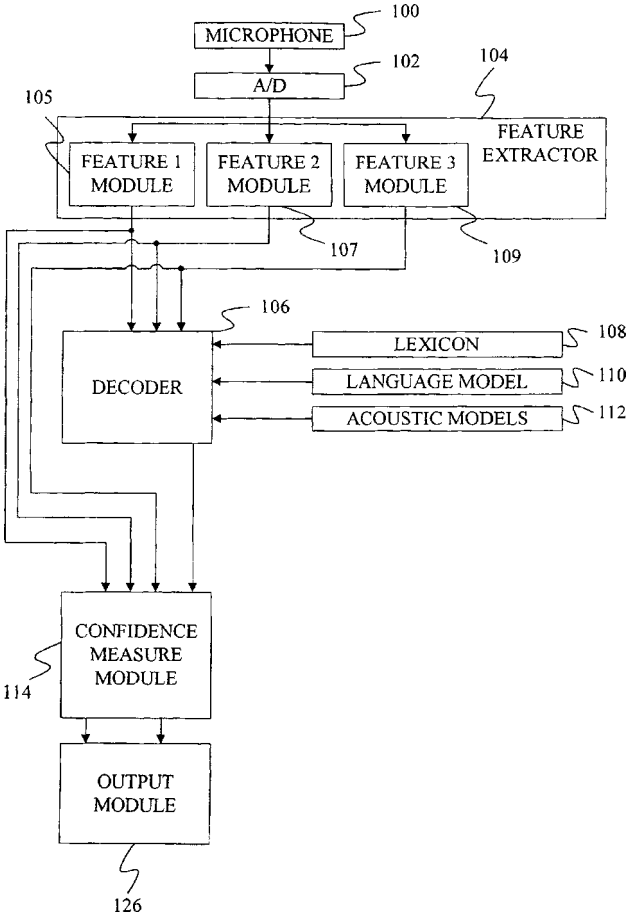
As described earlier in United States Patent 6,493,687 [reviewed in J. Acoust. Soc. Am. **113**(6), 2962 (2003)], an acoustic alarm system based on the signature of glass breaking can offer a simpler technique than attaching sensors to each window. However, when the glass is laminated the signature is different. This invention describes an analog chip architecture to detect such events.—MK

6,542,866

43.60.Lq SPEECH RECOGNITION METHOD AND APPARATUS UTILIZING MULTIPLE FEATURE STREAMS

Li Jiang and Xuedong Huang, assignors to Microsoft Corporation
1 April 2003 (Class 704/255); filed 22 September 1999

A speech signal is decoded based on at least two feature vectors per



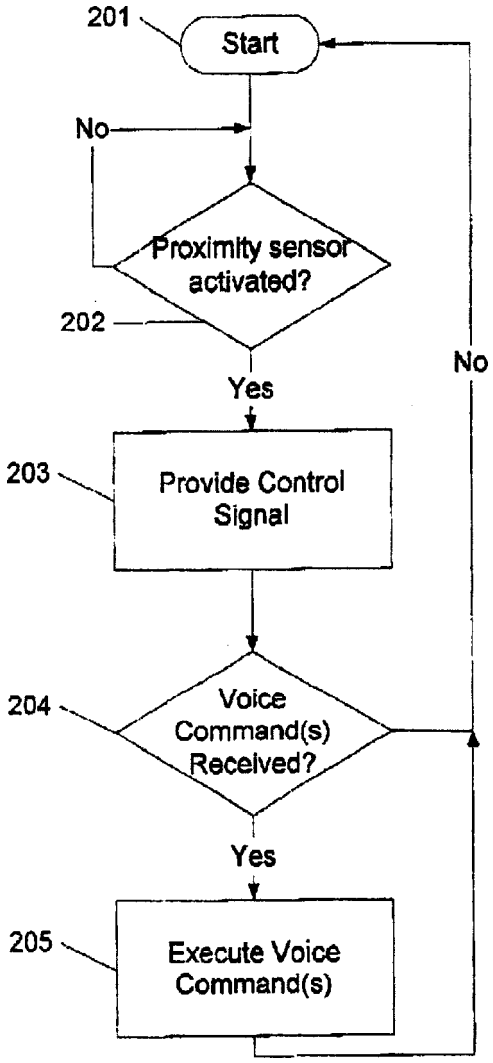
segment. The phonetic class membership probability of each segment is generated by selecting the best feature vector.—HHN

6,532,447

43.60.Qv APPARATUS AND METHOD OF CONTROLLING A VOICE CONTROLLED OPERATION

Nils Christensson, assignor to Telefonaktiebolaget LM Ericsson
11 March 2003 (Class 704/275); filed in Sweden 7 June 1999

A proximity sensor enables the voice recognition module of a mobile



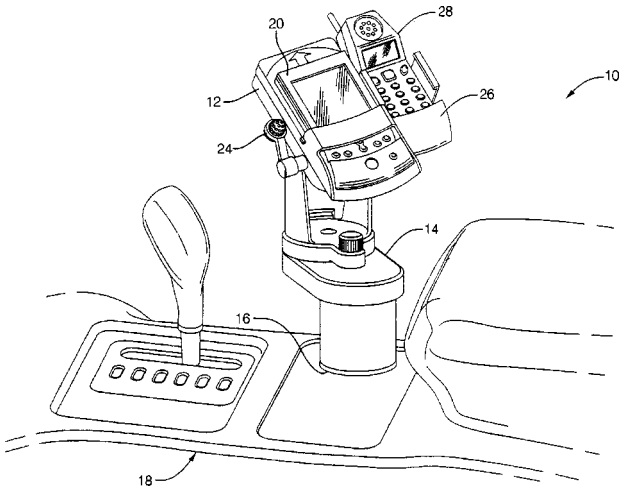
telephone device in noisy environments.—HHN

6,539,358

43.60.Qv VOICE-INTERACTIVE DOCKING STATION FOR A PROTABLE COMPUTING DEVICE

Bradley S. Coon and Ronald K. Reger, assignors to Delphi Technologies, Incorporated
25 March 2003 (Class 704/275); filed 24 May 2000

A voice-interactive docking station is provided by a computer device that includes at least one information management application, an associated



database, a speech recognizer, an interface, and a text-to-speech synthesizer.—HHN

6,535,853

43.64.Ri SYSTEM AND METHOD FOR DYSLEXIA DETECTION BY ANALYZING SPOKEN AND WRITTEN WORDS

Carmen T. Reitano, Salisbury, Massachusetts
18 March 2003 (Class 704/271); filed 14 August 2002

This dyslexia detection system displays a test word, either by itself or as part of a sentence. The test subject is requested to pronounce the test word and to write it out on a handwriting input tablet. The system analyzes the spoken words, comparing them to a stored record of the test word properly pronounced. The characters written by the test subject are also analyzed and the system detects whether any of the characters correspond to a member of a distorted character set containing distortions commonly produced by dyslexics, e.g., ''dear'' vs ''bear.'' A summary of mispronounced phonemes and writing errors is displayed.—DRR

6,535,131

43.64.Yp DEVICE AND METHOD FOR AUTOMATIC IDENTIFICATION OF SOUND PATTERNS MADE BY ANIMALS

Avshalom Bar-Shalom, Kibbutz Nachsholim and Gideon Inzler, Zichron Ya'acov, both of Israel
18 March 2003 (Class 340/573.1); filed in Israel 26 August 1998

The object of this device is to automatically identify sound patterns accompanying the activities and physiological processes of an animal, particularly sounds uttered by an animal in distress, and to alert an attendant at another location when help is needed. The device receives a sound pattern produced by the animal and attempts to match it with prestored audio patterns corresponding to sounds expected to be produced by that type of animal when in various types of distress (e.g., a pregnant animal going into labor). When a sufficiently good match is made, a signal is automatically relayed to a remote communication unit near an attendant.—DRR

6,457,362

43.70.Dn METHOD AND APPARATUS FOR DIAGNOSING AND REMEDIATING LANGUAGE-BASED LEARNING IMPAIRMENTS

Beverly A. Wright *et al.*, assignors to Scientific Learning Corporation
1 October 2002 (Class 73/585); filed 20 December 2001

This patent is another in a very long series assigned to the same company and covering various devices to help in retraining individuals with

certain phonetic perception deficits. The patents invariably refer to a malady known as specific language impairment (SLI). However, in general, SLI covers a much broader range of language learning problems than the narrow phonetic issues addressed here. The number of patients amenable to improvement using these techniques may not be as large as implied in the patent. This particular patent covers a system of masking intended to improve the subject's perception of consonant transitions.—DLR

6,463,406

43.72.Ar FRACTIONAL PITCH METHOD

Alan V. McCree, assignor to Texas Instruments Incorporated
8 October 2002 (Class 704/207); filed 20 May 1996

With original application in 1994, the subject matter of this patent was already well known at that time. The method used here for pitch resolution improvement is 16-times upsampling prior to the pitch analysis. A correlation is searched from a starting point taken from the low-rate analysis. The result can indeed be significant when starting from a rate of 8 K samples per second, especially for a high-pitched voice.—DLR

6,459,942

43.72.Kb ACOUSTIC COUPLING COMPENSATION FOR A SPEAKERPHONE OF A SYSTEM

Mitchell A. Markow *et al.*, assignors to Compaq Information Technologies Group, L.P.
1 October 2002 (Class 700/94); filed 30 September 1997

Manufactured speaker phones have become well-tuned devices capable of providing loudspeaker operation with nearly handset-like performance. It is a different matter when someone plugs a laptop with a mic and speaker into a phone jack. This patent offers a starting point for dealing with that issue. An equalization filter, typically implemented in laptop software, adapts so as to adjust the mic signal spectrum before that signal is affected by line echo processing.—DLR

6,542,582

43.72.Kb FILTERING SIGNALS IN A CONFERENCE CALL ENVIRONMENT

James O. Smith, Jr., assignor to WorldCom, Incorporated
1 April 2003 (Class 379/52); filed 24 July 2001

This is essentially the customary method for providing telephone services for the hearing-impaired that includes establishing a conference call between a hearing-impaired party, a hearing party, and a communications assistant (CA). The hearing party sends a voice message to the CA and the CA converts the voice message into a corresponding text message that is then relayed to the hearing-impaired party. The text message is displayed to the hearing-impaired party using a device that includes a modem. The modem produces a synchronization tone in response to receiving the text message. This potentially annoying synchronization tone and other tones are filtered to ensure that they do not reach the CA or the hearing party.—DRR

6,459,910

43.72.Ne USE OF SPEECH RECOGNITION IN PAGER AND MOBILE TELEPHONE APPLICATIONS

Theodore W. Houston, assignor to Texas Instruments Incorporated
1 October 2002 (Class 455/563); filed 7 June 1995

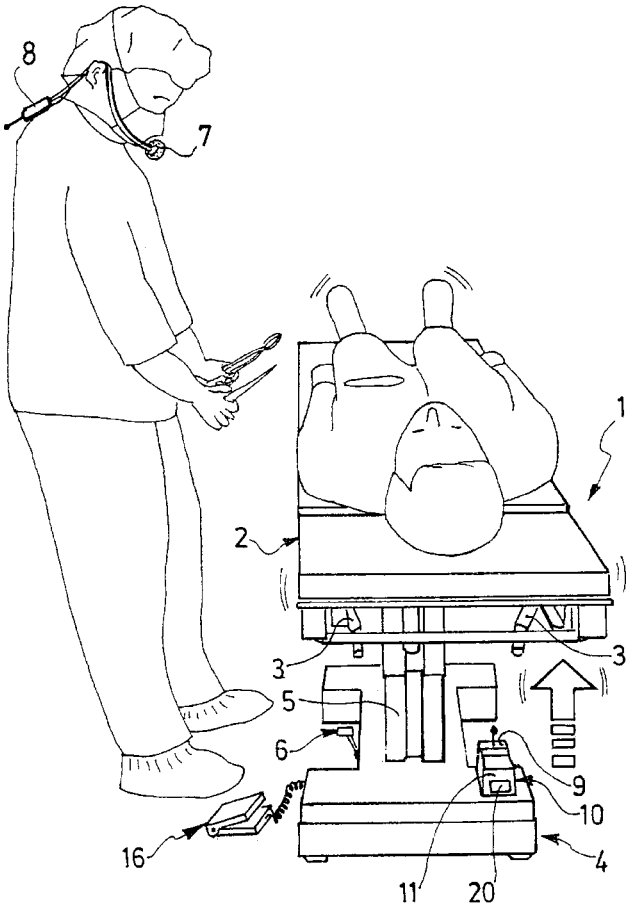
In a couple of pages, this patent covers the idea that a cell phone could use speech recognition algorithms as a form of compression. High quality recognition is not required. The idea is that any patterns that can be extracted and transmitted in a coded form serve to reduce the overall bandwidth. Examples include speaker gender, pitch, certains words or phrases, etc. Additional battery power required to operate the recognition software is not mentioned.—DLR

6,462,500

43.72.Ne OPERATING TABLE CONTROL SYSTEM AND OPERATING TABLE COMPRISING SUCH A SYSTEM

Jean-Marie L'Hegarat *et al.*, assignors to Alm
8 October 2002 (Class 318/649); filed in France 27 April 2000

This speech recognition application would allow a doctor to adjust the position of an operating table by voice commands. The authors believe that



a trained, small-vocabulary, speaker-dependent recognizer will provide adequate performance such that the doctor will suffer the headset rather than use the backup foot-pedal controls. Few other details are provided.—DLR

6,462,616

43.72.Ne EMBEDDED PHONETIC SUPPORT AND TTS PLAY BUTTON IN A CONTACTS DATABASE

Antony Paul Beswick and Mats Persson, assignors to Ericsson Incorporated
8 October 2002 (Class 329/88.03); filed 24 September 1998

The text of this patent acknowledges the difficulties encountered in using a recognition system for voice dialing. To be successful, such a system needs a database of names and phone numbers, preferably including phonetic information. On the other hand, most users will resist constructing a new database just for that purpose. This system seeks a middle ground, allowing the recognizer access to the user's existing phone number list, but also allowing the recognizer to insert phonetic information into the list when that information becomes available.—DLR

6,463,361

43.72.Ne SPEECH INTERFACE FOR AN AUTOMATED ENDOSCOPIC SYSTEM

Yulun Wang and Darrin Uecker, assignors to Computer Motion, Incorporated
8 October 2002 (Class 700/258); filed 22 September 1994

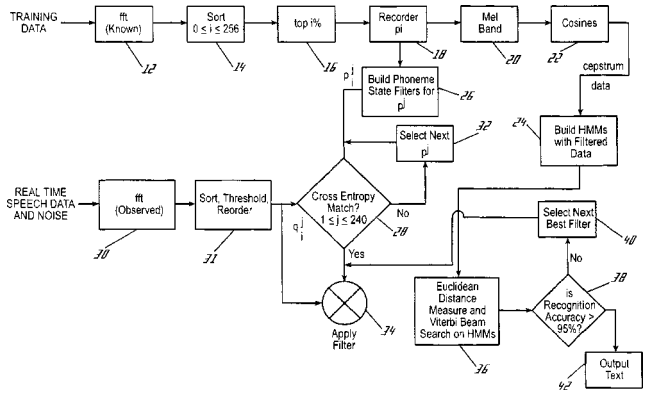
The patent describes a robotic medical assistant, such as a computer-controlled endoscope, which is also provided with a speech recognizer and a voice command architecture. As a safety feature, the user would have to speak a specific keyword at the beginning of each command. The specific commercial recognizer suggested in the patent is no longer manufactured and, in fact, the company named has been twice swallowed by larger companies. Most of the patent text deals with the geometry of the robotic arm movements.—DLR

6,529,866

43.72.Ne SPEECH RECOGNITION SYSTEM AND ASSOCIATED METHODS

R. Bradley Cope and Stephen G. Boemler, assignors to The United States of America as represented by the Secretary of the Navy
4 March 2003 (Class 704/205); filed 24 November 1999

Filtered noisy speech signals are compared to a predetermined model



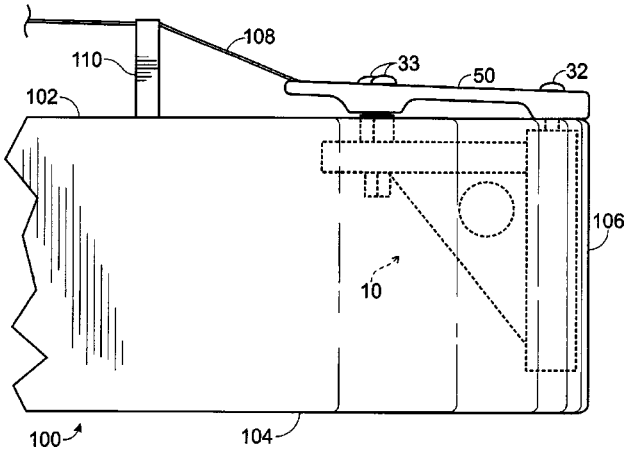
of the speech signal. The closest match is selected and displayed.—HHN

6,534,697

43.75.Gh CANTILEVERED TAIL BLOCK/TAILPIECE FOR AN ACOUSTIC GUITAR

Kevin K. Ko, Eugene, Oregon
18 March 2003 (Class 84/299); filed 16 May 2002

Given the stress created by the string tension, guitars and other



stringed instruments are susceptible to breakage and or bending. The inventor proposes adding a cantilevered tail as shown.—MK

6,538,183

43.75.Gh COMPOSITE STRINGED MUSICAL INSTRUMENT, AND METHOD OF MAKING THE SAME

Frederick J. Verd, Santee, California
25 March 2003 (Class 84/291); filed 7 February 2001

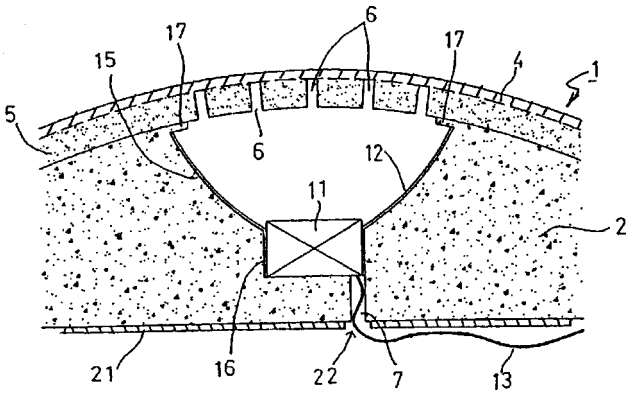
The bulk density/modulus of a material is an important aspect of an instrument's sound quality. The question here is how to create a totally artificial acoustic stringed instrument such as a guitar. Accordingly, the inventor proposes three layers: a matrix epoxy on the outside with carbon fiber reinforcement and an elastomeric inner layer. As with any layer cake, the proof is in the eating. The inventor claims ''exceptionally pleasing sound quality.''—MK

6,527,559

43.80.Qf HUMAN SIZED MANIKIN FOR TRAINING OF AUSCULTATION

Chiharu Yoshii *et al.*, assignors to Kyoto Kagaku Company, Limited
4 March 2003 (Class 434/266); filed in Japan 31 October 2000

Human-sized manikins used for training in auscultation typically have acoustical characteristics not quite like those of a human body. In order to overcome the problems of sound attenuation by intervening materials which tend to mask acoustic subtleties, a design is proposed for a manikin that uses loudspeakers coupled to a body sound reproducing apparatus and is covered



by an imitation skin **1** with an inner cover **5** on the outside of the urethane foam core body **2**. Recesses **15** are incorporated into the core body **2** to correspond to sites to be auscultated. A concave surface **12**, attached to each speaker **11**, serves as a reflector. Each speaker-reflector combination is nested into each recess **15** and supported by the core.—DRR

6,527,717

43.80.Qf TISSUE MOTION ANALYSIS MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD

John I. Jackson *et al.*, assignors to Acuson Corporation
4 March 2003 (Class 600/437); filed 30 June 2000

With this method, the motion of the ultrasound transducer establishes tissue motion. Correction for transducer motion aids in isolating localized tissue contractions or expansions, e.g., the motion of myocardial muscle or fibers. Determining an angle of motion from the ultrasound data supposedly provides accurate estimates of motion. The angle of motion is used to adjust velocity estimates, thereby providing two-dimensional velocity vectors. Movement of the tissue is established by correlating a distinguishing feature such as a speckle represented by two different sets of ultrasound data obtained at different times.—DRR

6,527,721

43.80.Qf PORTABLE ULTRASOUND SYSTEM WITH BATTERY BACKUP FOR EFFICIENT SHUTDOWN AND RESTART

Paul Wittrock *et al.*, assignors to Koninklijke Philips Electronics, N.V.
4 March 2003 (Class 600/446); filed 24 October 2000

This portable ultrasound system is designed to be turned off quickly either by operator command or in response to an interruption of ac power to the system. When the ac power is cut off, the system switches to battery backup so that a processor can execute an orderly shutdown sequence. The system can be shut down completely, or the state of the system can be minimally preserved through battery backup in either volatile or nonvolatile memory so that the system can restart without going through a full boot-up sequence. Thus the system can remain active even when it is unplugged and being moved.—DRR

6,527,723

43.80.Qf VARIABLE MULTI-DIMENSIONAL APODIZATION CONTROL FOR ULTRASONIC TRANSDUCERS

William J. Ossmann and McKee D. Poland, assignors to Koninklijke Philips Electronics N.V.
4 March 2003 (Class 600/459); filed 26 June 2001

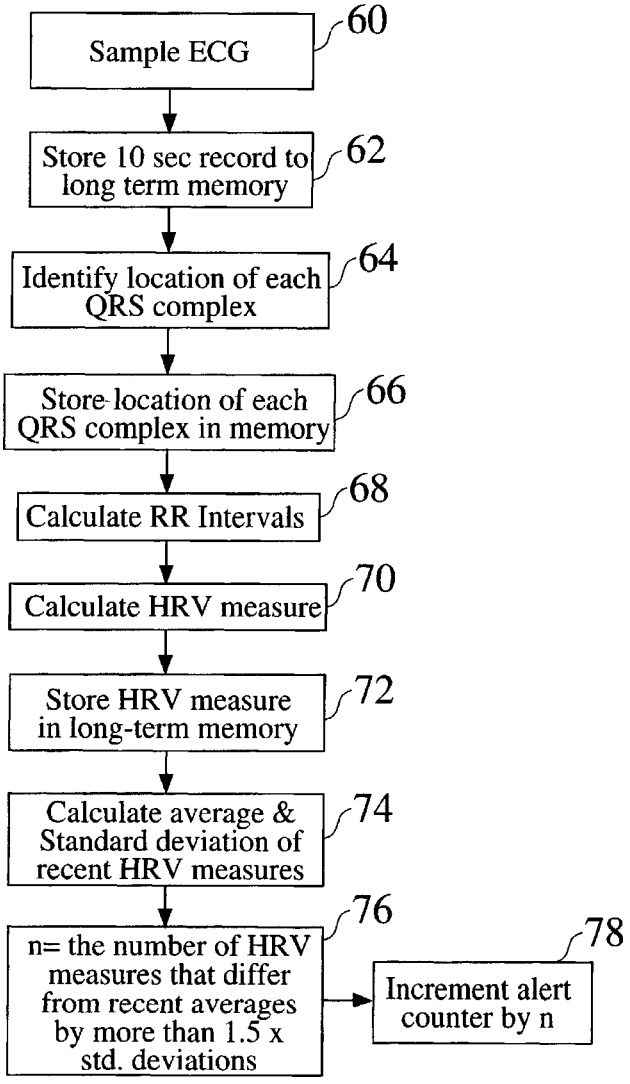
For medical imaging, it is desirable to have variable multi-dimensional control of an ultrasonic transducer array that allows all of its dimensions to have variable apodization control. The control described in this patent applies to both piezoelectric-based transducers and to micromachined ultrasonic transducers. The transmit-receive capabilities of the transducers are used to provide feedback to a controller programmed to provide an apodization profile.—DRR

6,527,729

43.80.Qf METHOD FOR MONITORING PATIENT USING ACOUSTIC SENSOR

Robert Turcott, assignor to Pacesetter, Incorporated
4 March 2003 (Class 600/528); filed 11 October 2000

This system monitors the progression of the disease of a heart failure patient. An implantable or other ambulatory monitor senses acoustic signals from the patient's lungs and heart. Significant changes in the energy content



of either heart or lung sounds generally denote exacerbation of heart failure. The information may be used to alert the patient or medical attendants to changes in the patient's condition that may warrant attention.—DRR

6,530,887

43.80.Qf ULTRASOUND PROBE WITH INTEGRATED ELECTRONICS

Jeffrey M. Gilbert *et al.*, assignors to TeraTech Corporation
11 March 2003 (Class 600/459); filed 26 November 1999

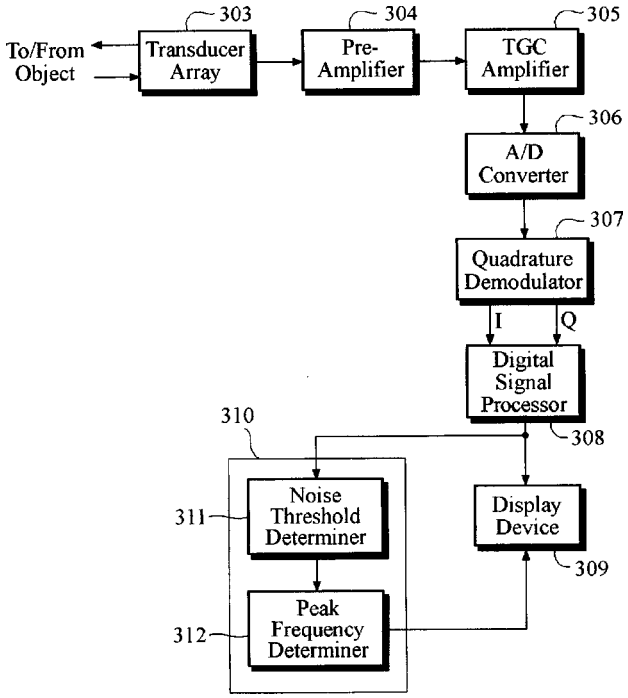
The idea of this hand-held ultrasonic probe is to integrate control, beam forming, and transducer drive circuitry within an ergonomic housing instead of accommodating these components in a nearby console. The probe electronics communicates with a host computer using a standard high-speed serial bus.—DRR

6,530,890

43.80.Qf ULTRASOUND DIAGNOSTIC APPARATUS AND METHOD FOR MEASURING BLOOD FLOW VELOCITY USING DOPPLER EFFECT

Ji Hoon Bang and Cheol An Kim, assignors to Medison Company, Limited
11 March 2003 (Class 600/504); filed in the Republic of Korea 8 July 2000

This ultrasound diagnostic device measures blood flow velocity by effectively selecting either the positive or the negative frequency range to compute a noise threshold and to determine its reliability. A sample data generator samples a reflection of an ultrasound signal that was transmitted into the body. The frequency distribution indicates a number of frequency components. Either a positive or a negative frequency distribution range is



selected and a noise threshold is established by using a predetermined number of frequency components within the selected frequency range. The peak frequency component is determined which has the highest frequency among the frequency components. Each frequency component corresponds to a power level higher than the noise threshold, and the peak frequency component corresponds to the peak blood frequency velocity.—DRR

6,533,736

43.80.Qf WIRELESS MEDICAL STETHOSCOPE

Mark Moore, Tallahassee, Florida
18 March 2003 (Class 600/586); filed 30 May 2000

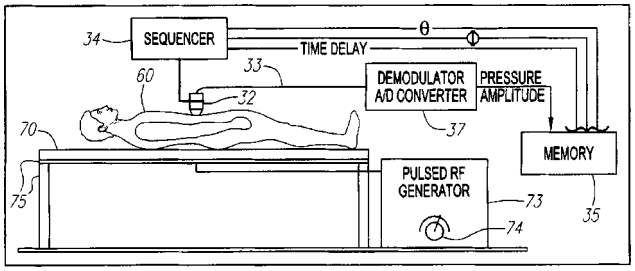
This wireless stethoscope consists of two parts: an auscultation piece and a hearing piece. A microphone and a conventional rf chip are housed within the auscultation piece to enable rf transmissions. The hearing piece contains a link manager receiver device to accept the radio signals from the transmission system and a means of converting the radio signals into sound.—DRR

6,535,625

43.80.Qf MAGNETO-ACOUSTIC IMAGING

David B. Chang *et al.*, assignors to Magnetus LLC
18 March 2003 (Class 382/128); filed 24 September 1999

The apparatus consists of a coil of wire **32** that is energized by a pulsed rf power source to generate electromagnetic fields within a body. A scanning, directional hydrophone acoustically attached to the body samples ultrasonic radiation generated by the rf field at or twice the rf frequency



induced from conductive spots. The signals from the hydrophone are analyzed with respect to the time and direction of arrival of the rf pulses. The collected data are recorded and displayed as images of the spots within the region.—DRR

6,537,223

43.80.Qf PROBE HEAD

Knut Kristiansen, assignor to Medi-Stim AS
25 March 2003 (Class 600/459); filed 5 September 2000

This probe head is a one-piece sealed ultrasound device for measuring flow in vessels. It consists of a conventional ultrasound reflector and piezoelectric crystals mounted on a support element that is permanently inserted into a housing fabricated from resilient material. It also includes gel spill guards for limiting acoustic coupling gel spillage and vessel retainer flaps to retain the vessel in a stationary position in the probe head.—DRR

6,537,233

43.80.Qf AUDITORY DISPLAY OF KNEE JOINT VIBRATION SIGNALS

Rangaraj M. Rangayyan *et al.*, assignors to University Technologies International Incorporated
25 March 2003 (Class 600/586); filed 6 November 2000

Audification and sonification are used in computer-assisted auscultation of knee joints to yield auditory display of vibroarthrographic signals emitted during active movement of the leg. In audification, the vibroarthrographic signals are scaled in time and frequency using a time-frequency

distribution to facilitate aural analysis. In sonification, the instantaneous mean frequency and envelope of the vibroarthrographic signals are derived and used to synthesize sounds that can facilitate more accurate diagnosis.—DRR

6,537,234

43.80.Qf VIBRATION TRANSDUCER

Akira Komatsu, assignor to Acouve Laboratory, Incorporated
25 March 2003 (Class 601/46); filed in Japan 28 October 1998

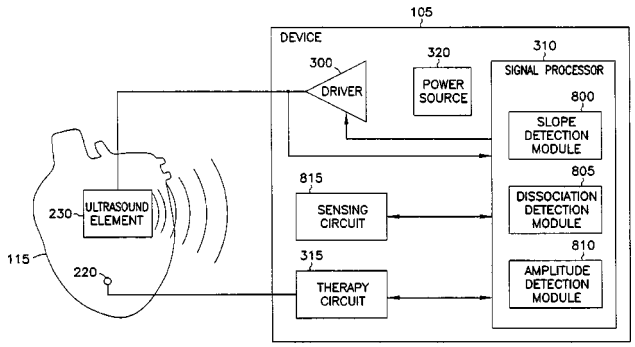
This patent refers to the structure of an electromechanical vibration transducer used as apparatus for producing body-sensible sound vibration by an electric signal in the low frequency range.—DRR

6,539,262

43.80.Qf CARDIAC RHYTHM MANAGEMENT SYSTEM WITH ULTRASOUND FOR AUTOCAPTURE OR OTHER APPLICATIONS

Robert J. Sweeney, assignor to Cardiac Pacemakers, Incorporated
25 March 2003 (Class 607/28); filed 10 September 2001

This cardiac pacemaker provides ultrasound autocapture capability for determining whether stimulation has evoked a desired response from the heart. It also provides for adjusting the stimulation energy on the basis of the observed response from the heart. An ultrasound element is deployed on a lead in the heart. Another ultrasound element is located elsewhere in the



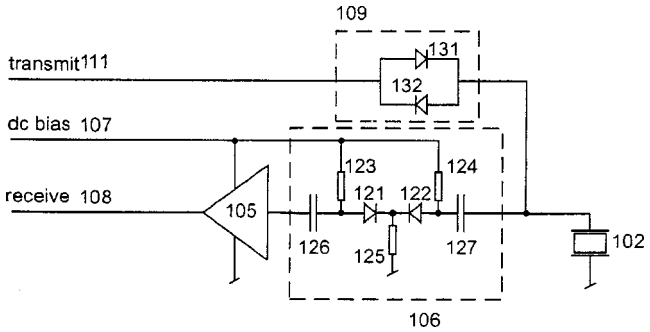
heart or in the implant device. An autocapture determination circuit establishes whether the motion of the heart chamber indicates a contraction in response to the stimulation and adjusts the stimulation energy to provide only the energy necessary to provide capture. This is said to save energy, prolonging the life of the implanted device.—DRR

6,540,677

43.80.Qf ULTRASOUND TRANSCEIVER SYSTEM FOR REMOTE OPERATION THROUGH A MINIMAL NUMBER OF CONNECTING WIRES

Bjorn A. J. Angelsen and Tonni F. Johansen, both of Trondheim, Norway
1 April 2003 (Class 600/437); filed 15 November 2001

This is a device that relates to high resolution ultrasound imaging of small structures at high frequencies, typically above 5 MHz, where the ultrasound transducer (or transducer array) is brought close to the structure to be imaged through channels with limited space. An example of such



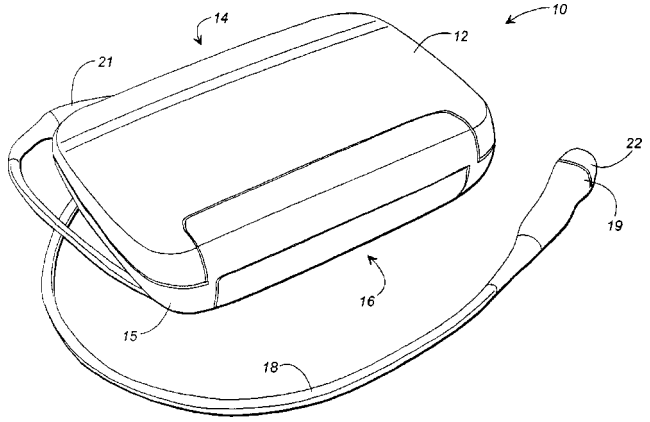
applications is intravascular ultrasound imaging of a vessel wall from a transducer on the tip of a catheter. In this design, integration of preamplifier electronics and circuits for acoustic beam forming with ultrasonic arrays allows for minimizing the number of connecting electric wires. A high-voltage transmit pulse shares the same cable with the low-level receive signal through the use of a ''transmit switch,'' shown as **109** in the figure.—DRR

6,540,685

43.80.Qf ULTRASOUND DIAGNOSTIC DEVICE

Peter Koester Rhoads *et al.*, assignors to Koninklijke Philips Electronics N.V.
1 April 2003 (Class 600/459); filed 9 November 2000

Miniaturization of an ultrasound diagnostic device results in an extremely compact and lightweight portable unit, smaller than a typical laptop computer. The unit consists of electrical control circuitry, a battery pack, a



carrying strap, and a transducer assembly. A display and a console are hinged together so as to form a single package that unsnaps to open, in a fashion similar to a laptop computer.—DRR

6,527,759

43.80.Sh ULTRASOUND ASSEMBLY FOR USE WITH LIGHT ACTIVATED DRUGS

Katsuro Tachibana *et al.*, assignors to Ekos Corporation
4 March 2003 (Class 604/500); filed in Japan 5 March 1995

A method and apparatus are disclosed for causing localized tissue destruction within a tissue site. The apparatus consists of a catheter with a lumen coupled with a media delivery port through which a light-activated drug can be locally delivered to the tissue site. The light activated drug becomes activated upon exposure to a particular level of ultrasound energy, owing to the sonoluminescent effect.—DRR

6,527,763

43.80.Sh FLOW APPARATUS FOR THE DISRUPTION OF OCCLUSIONS

Victor C. Esch *et al.*, assignors to Endovasix, Incorporated
4 March 2003 (Class 606/2); filed 6 September 2001

The device, intended to remove material forming a partial or total occlusion of a vessel, consists of a catheter that incorporates an ability to ''chew'' through an occlusion. Flow generated through an active distal portion of the catheter draws the occlusion toward a nearby optical fiber which is a source of acoustic pressure, shock waves, or other forces. The repetitive formation and collapse of bubbles in the vessel fluid helps in breaking up the occlusion, thereby giving rise to recanalization of the fluid flow within the vessel.—DRR

6,533,803

43.80.Sh WOUND TREATMENT METHOD AND DEVICE WITH COMBINATION OF ULTRASOUND AND LASER ENERGY

Eilaz Babaev, assignor to Advanced Medical Applications, Incorporated
18 March 2003 (Class 607/89); filed 22 December 2000

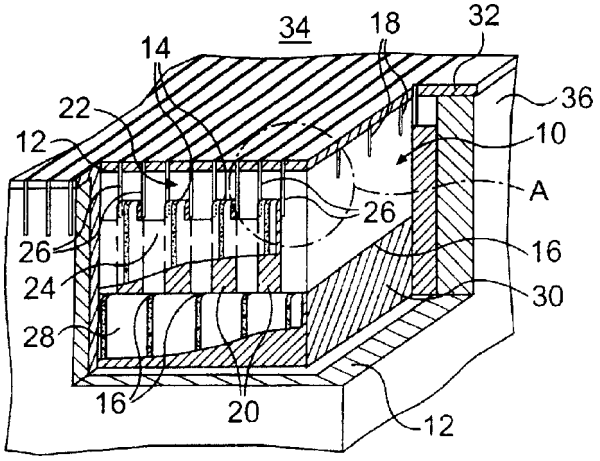
This device provides treatment of wounds through the use of ultrasound and laser energy. The method involves spraying liquid drugs onto the wound surface using ultrasonic waves applied through a transducer with a distal tip constituting the radiation surface. The system can deliver drugs, kill bacteria, cleanse a surface, and stimulate healthy tissue cells and the wound can be treated with laser energy. It is said that the combination of ultrasonic spraying and laser application results in a higher level of disinfection by the energized liquids as compared to liquids applied by ordinary methods.—DRR

6,537,224

43.80.Sh MULTI-PURPOSE ULTRASONIC SLOTTED ARRAY TRANSDUCER

Pascal Mauchamp and Aimé Flesch, assignors to Vermon
25 March 2003 (Class 600/459); filed 8 June 2001

This probe, intended for use in high resolution imaging for therapy or other high-intensity applications, contains two ultrasonic transducer arrays,



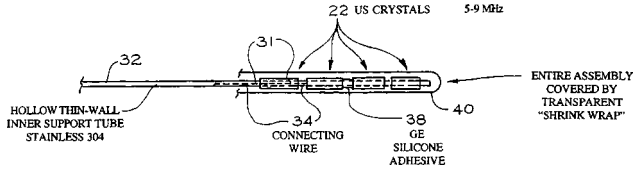
each operating at separate resonant frequencies and formed by a number of elements arranged linearly along a coordinate axis. The elements of the two arrays are interdigitized.—DRR

6,537,306

43.80.Sh METHOD OF MANUFACTURE OF A TRANSURETHRAL ULTRASOUND APPLICATOR FOR PROSTATE GLAND THERMAL THERAPY

Everette C. Burdette and Chris J. Diederich, assignors to The Regents of the University of California; Dornier Medical Systems, Incorporated
25 March 2003 (Class 607/96); filed 19 May 1997

The subject of the patent is a device for applying thermal energy to a prostate gland. It includes a tube with a longitudinal passageway and a wall with radial apertures. Power leads are threaded through the tube and out the radial apertures in the tube wall. Ultrasound crystals are distributed along



the tube and the power leads are coupled through the apertures to the ultrasound crystals. A sealant is coupled acoustically to the ultrasound crystals. It is said that the device can be used to shape the ultrasound treatment fields and to minimize application of energy to rectal tissue and the like.—DRR

6,540,700

43.80.Sh ULTRASOUND TREATMENT APPARATUS

Katsuhiko Fujimoto *et al.*, assignors to Kabushiki Kaisha Toshiba
1 April 2003 (Class 601/3); filed in Japan 26 October 1998

This apparatus provides ultrasound treatment for necrosing a tumor by focusing the ultrasound on the tumor. The system consists of an ultrasound source for generating focused ultrasound, a transducer driving circuit, and a controller. Ultrasound is generated under an irradiation condition in which an optimization index based on the product of the focus intensity (in W/cm^2), the irradiation period (in seconds), and the operating frequency (in MHz) falls within the range of from 6000 to 40 000.—DRR

6,527,718

43.80.Vj ULTRASOUND SYSTEM FOR CONTINUOUS IMAGING AND DELIVERY OF AN ENCAPSULATED AGENT

Brian G. Connor, Newfields, New Hampshire and Matthew Mooney, Westford, Massachusetts
4 March 2003 (Class 600/439); filed 28 March 2000

This system uses a transducer mounted within a housing that is secured to a patient. The sound field produced by the transducer causes the carrier of a drug to release the agent. The release of the agent is controlled by comparing an ultrasonic measurement of a physiological parameter to a threshold and adjusting the power of the sound field.—RCW

6,527,719

43.80.Vj ULTRASONIC DIAGNOSTIC IMAGING SYSTEM WITH REDUCED POWER CONSUMPTION AND HEAT GENERATION

Lars Jonas Olsson *et al.*, assignors to Koninklijke Philips Electronics N.V.
4 March 2003 (Class 600/443); filed 24 October 2000

When this system is idle for a period of time, a sequence of energy-saving procedures is initiated that includes preparing the system for use at a specified time.—RCW

6,527,720

43.80.Vj MEDICAL ULTRASONIC IMAGING METHOD AND SYSTEM FOR SPATIAL COMPOUNDING

Kutay F. Ustuner *et al.*, assignors to Acuson Corporation
4 March 2003 (Class 600/443); filed 24 September 2001

Improved detectability of targets that have a strong directional response as well as targets that produce speckle is obtained by compounding multiple images produced by using only one or two transmit pulses per ultrasonic b-scan line.—RCW

6,530,885

43.80.Vj SPATIALLY COMPOUNDED THREE DIMENSIONAL ULTRASONIC IMAGES

Robert R. Entrekin and James R. Jago, assignors to ATL Ultrasound, Incorporated
11 March 2003 (Class 600/437); filed 17 March 2000

An ultrasonic probe is moved to scan a volume. As the probe is moved, targets within the volume are interrogated from multiple directions. The echo data from these directions are spatially compounded and processed for display in a three-dimensional format.—RCW

6,530,888

43.80.Vj IMAGING PROBES AND CATHETERS FOR VOLUMETRIC INTRALUMINAL ULTRASOUND IMAGING

Stephen W. Smith and Edward D. Light, assignors to Duke University
11 March 2003 (Class 600/463); filed 27 July 2001

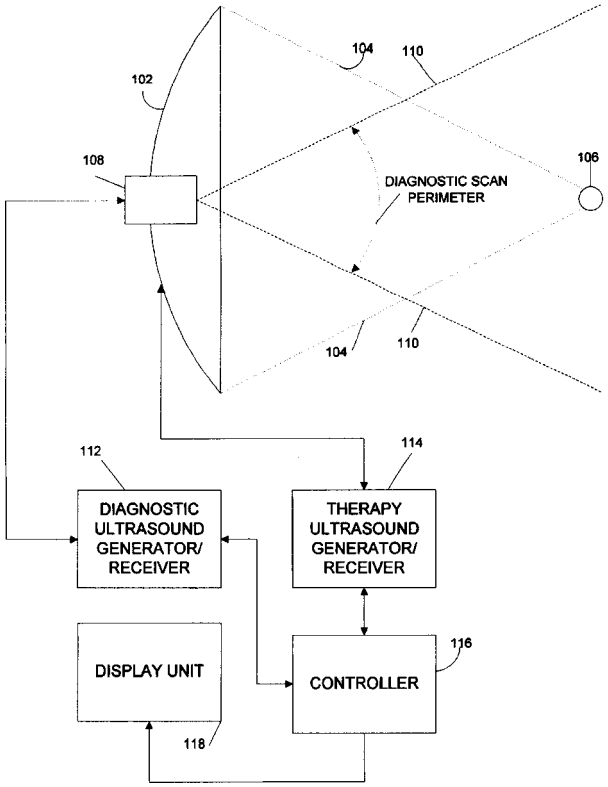
An ultrasonic phased array transducer is positioned to emit ultrasonic beams and receive echoes from a volume forward of the distal end of a catheter. From an interstitial site defined by the absence of one or more phased array elements, a tool such as a fiber optic lead, a suction device, a guide wire, or an ablation electrode is positioned. Images can be obtained from the region in front of the catheter tip or from a region around a perpendicular to the axis of the catheter.—RCW

6,533,726

43.80.Vj SYSTEM AND METHOD FOR ULTRASONIC HARMONIC IMAGING FOR THERAPY GUIDANCE AND MONITORING

Frederic Louis Lizzi and Cheri Xiaoyu Deng, assignors to Riverside Research Institute
18 March 2003 (Class 600/439); filed 8 August 2000

A diagnostic ultrasound imaging transducer is mounted with an ultrasonic therapy transducer to produce images of a region receiving therapy. The difference between images obtained before and after a stage in the



therapy is obtained by subtracting the images to show changes resulting from the therapy. Progress in therapy is also monitored by evaluating transient changes due to heating as well as changes from alteration of cell microstructure.—RCW

6,533,727

43.80.Vj ULTRASONIC SYSTEM AND METHOD EMPLOYING NON-INTEGER HARMONIC ECHO SIGNALS FOR IMAGING

George A. Brock-Fisher, assignor to Koninklijke Phillips Electronics N.V.
18 March 2003 (Class 600/458); filed 11 February 2000

Ultrasound signals are emitted in a band around a fundamental frequency and echoes resulting from interaction of the signals with both tissue and a microbubble contrast agent are received. The power level of the transmitted signal is controlled to destroy the microbubbles. The echoes are processed to selectively extract signals at a noninteger multiple of the fundamental frequency. Images are produced principally from the extracted signal components.—RCW

6,533,728

43.80.Vj METHOD AND APPARATUS FOR RECOVERY AND PARAMETRIC DISPLAY OF CONTRAST AGENTS IN ULTRASOUND IMAGING

Marek Belohlavek and Richard Y. Bae, assignors to Mayo Foundation for Medical Education and Research
18 March 2003 (Class 600/458); filed 20 November 2001

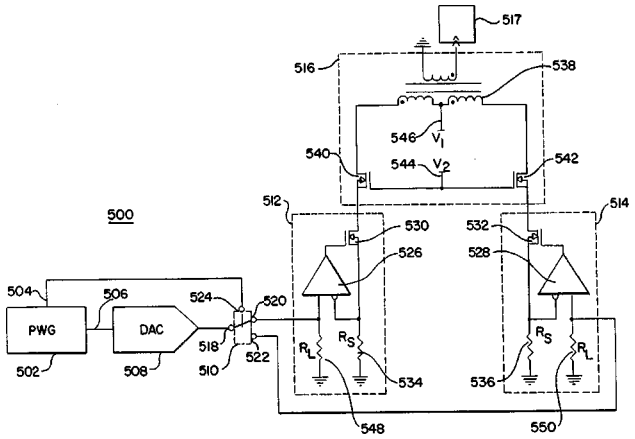
Echo signals along a scan line are processed to obtain spectral components around a fundamental frequency and also around a harmonic frequency. A comparison of the peak amplitudes of the two spectral components determines the presence of a contrast agent. The comparison is used to indicate perfusion.—RCW

6,537,216

43.80.Vj TRANSMIT CIRCUIT FOR IMAGING WITH ULTRASOUND

Lazar A. Shifrin, assignor to Acuson Corporation
25 March 2003 (Class 600/437); filed 30 April 2001

A digital-to-analog converter producing a unipolar waveform that contains positive and negative portions of a desired bipolar waveform is connected to a switch. The switch selects current driver paths as a function of



the positive or negative polarity of the desired bipolar waveform. The current drivers feed a push–pull output amplifier that produces the desired bipolar excitation waveform.—RCW

6,537,217

43.80.Vj METHOD AND APPARATUS FOR IMPROVED SPATIAL AND TEMPORAL RESOLUTION IN ULTRASOUND IMAGING

Steinar Bjærum *et al.*, assignors to GE Medical Systems Global Technology Company, LLC
25 March 2003 (Class 600/441); filed 24 August 2001

A set of ultrasound pulses is transmitted at a frame rate in one mode of operation and the echoes are received. Another set of ultrasound pulses is transmitted at a different frame rate using another mode of operation. The first set of pulses defines a range, while the second set defines a subrange. Echoes from the second set of pulses are received and the echoes from the first and second sets are displayed as a single image.—RCW

6,537,218

43.80.Vj MEDICAL DIAGNOSTIC IMAGING SYSTEM AND METHOD FOR ADAPTIVE INTERLINE DELAY AND TRANSMIT POWER

Constantine Simopoulos *et al.*, assignors to Acuson Corporation
25 March 2003 (Class 600/447); filed 25 April 2000

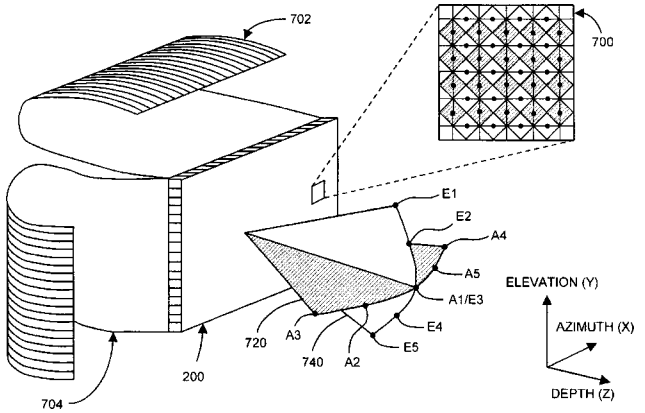
Either the interval between transmit pulses or the transmit power is set adaptively based on an analysis of echoes to maintain a high frame rate without wraparound artifacts associated with transmit pulses that are too closely spaced.—RCW

6,537,220

43.80.Vj ULTRASOUND IMAGING WITH ACQUISITION OF IMAGING DATA IN PERPENDICULAR SCAN PLANES

Barry Hugh Friemel and Charles Dale Emery, assignors to Siemens Medical Solutions USA, Incorporated
25 March 2003 (Class 600/447); filed 31 August 2001

Two independent, interleaved linear subarrays that can be steered independently and focussed in different imaging planes form the face of a probe.



By transmitting in one plane and receiving in an orthogonal plane, a three-dimensional region of interest can be imaged. Two orthogonal b-scans of a structure can be displayed simultaneously.—RCW

6,537,221

43.80.Vj STRAIN RATE ANALYSIS IN ULTRASONIC DIAGNOSTIC IMAGES

Aline Laure Criton *et al.*, assignors to Koninklijke Philips Electronics, N.V.
25 March 2003 (Class 600/454); filed 7 December 2000

The spatial gradient of velocity in the direction of tissue motion is used to calculate strain rate. The direction of motion in cardiac ultrasound images can be either in the plane of the myocardium or across the myocardium. Strain rate information calculated for a sequence of images is displayed over the full heart cycle for an automatically drawn outline such as the endocardial border.—RCW

6,540,679

43.80.Vj VISUAL IMAGING SYSTEM FOR ULTRASONIC PROBE

Michael H. Slayton and Peter G. Barthe, assignors to Guided Therapy Systems, Incorporated
1 April 2003 (Class 600/439); filed 28 December 2000

This system obtains an image of the ultrasonic transducer position while the transducer is being used for diagnosis or therapy. An ultrasonic image produced by the acoustic probe can be combined with the corresponding image of the transducer position to show the location of the transducer with respect to a patient and the ultrasonic image of the region being scanned. Information about the transducer position on a patient can be used to capture similar imaging planes on the patient or from subsequent patients.—RCW

6,540,681

43.80.Vj EXTENDED VIEW ULTRASOUND IMAGING SYSTEM

Xiangyong Cheng and Shengtz Lin, assignors to U-Systems, Incorporated
1 April 2003 (Class 600/443); filed 24 November 2000

Successive frames of a scan along with the location and orientation of the frames are processed to produce an extended view of a target region. The field of the extended view is first initialized and then updated using information from a sensor that detects the location and orientation of the transducer during each frame. A weighting factor that may be adjusted by the user is used to weight the current frame in the composite image. The weighting factor may be automatically or dynamically adjusted by using the reflectivity of structures, edge motion, or time in a periodic cycle.—RCW

6,540,682

43.80.Vj PORTABLE, CONFIGURABLE AND SCALABLE ULTRASOUND IMAGING SYSTEM

Steven C. Leavitt *et al.*, assignors to Koninklijke Philips Electronics N.V.
1 April 2003 (Class 600/447); filed 9 November 2000

This system consists of a phased ultrasonic transducer array connected to a processor. When a sector array is employed, the system uses a processing channel associated with each element in the array to develop the ultrasonic image. When a linear or curved transducer array is employed, the system uses fewer processing channels to develop the image. The system has a scaleable architecture and includes software for various imaging applications, operating modes, and processing algorithms.—RCW

6,540,683

43.80.Vj DUAL-FREQUENCY ULTRASONIC ARRAY TRANSDUCER AND METHOD OF HARMONIC IMAGING

Gregory Sharat Lin, Fremont, California
1 April 2003 (Class 600/447); filed 14 September 2001

This transducer is a linear array of alternating long and short elements. One set of elements is used to transmit and receive in a band around a fundamental frequency and a second set of elements is used to receive in a band around the second harmonic of the fundamental frequency or around subharmonic bands. Transmit beamforming is performed at the fundamental frequency and receive beamforming is done at the second harmonic or subharmonic frequency. When receive beamforming at the fundamental frequency is included, images at the fundamental frequency or harmonic frequency can be obtained and images that are the sum or difference of images at the fundamental and harmonic frequencies can also be obtained.—RCW

6,540,684

43.80.Vj ULTRASONIC PERFUSION MEASUREMENT USING CONTRAST AGENTS

Michalakis Averkiou *et al.*, assignors to Koninklijke Philips Electronics N.V.
1 April 2003 (Class 600/458); filed 8 November 2001

A real time image showing the location of a contrast agent and a triggered contrast image are shown by using various methods of contrast agent detection and imaging.—RCW

Acoustical losses in wet instrument bores (L)

John W. Coltman
3319 Scathelocke Road, Pittsburgh, Pennsylvania 15235

(Received 16 May 2003; revised 25 June 2003; accepted 7 July 2003)

The possibility that condensed water in wind instrument bores might contribute to acoustic losses through evaporation and condensation during the acoustic cycle was explored in a short experiment. A wet tube exhibited losses about 4% higher than when dry, the precision of the measurements being 0.6 percentage points, strongly suggesting that this mechanism is operative. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1603766]

PACS numbers: 43.20.Mv [NHF]

All wind instruments operate under conditions where the most of the air inside the tube is saturated with water, and liquid water is present. Nevertheless, laboratory measurements are usually conducted dry, with air at relative humidity typical of the ambient atmosphere. In a conversation, Peter Hoekje of Baldwin-Wallace College pointed out that water in the tube could evaporate during the negative pressure portion of the standing wave and condense again when the pressure is positive, and that this process might contribute to the acoustical losses in the bore.

Having never seen any allusion to such a process in the literature, it seemed worthwhile to undertake a simple experiment to see if such an effect were present. The experiment described here gave evidence that the effect, though small compared to other loss mechanisms, probably exists.

A copper pipe 93 cm long with inside diameter 19 mm was closed at one end by a piston driver and at the other by a rigid ceramic microphone. The first two resonance modes of the system were at 186 and 374 Hz. The Q's of these resonances were evaluated by fitting measured responses around the resonance frequency to a theoretical resonance curve. The Q's obtained were consistent with previously reported[1] attenuation coefficients for this size tube.

First, the tube, whose inside surface was polished, was swabbed with alcohol, allowed to dry, and sealed in the measuring system with a small sack of silica gel to act as a dryer. The sack was left in overnight, and then removed before the Q measurement was made, care being taken to admit the least possible room air into the tube. After the measurement, the tube was wetted inside, and the measurement repeated. The first wetting was performed by breathing a number of times into the tube, until a haze could be seen on the wall, and then sealing the tube. A second technique was to spray the inside with a flower mister, and shake out any drops that did not adhere to the walls. Numerous small droplets adhering to the walls were visible. Measurements showed no significant dependence on the method of wetting. The process of measuring dry and then wet was repeated three times over a few days.

Measured values of Q for the first mode at 186 Hz were 36.48 ± 0.13 when dry and 35.25 ± 0.22 when wet. For the second mode at 374 Hz the values were 53.00 ± 0.15 and 50.65 ± 0.11, respectively. These correspond to attenuation coefficients of 0.0463/m and 0.0637/m for the first and second modes when dry.

According to these measurements wetting the tube increased the losses by 3.5% at 186 Hz and 4.6% at 374 Hz. The standard deviations on these numbers are 0.7 and 0.4 percentage points, respectively. These were derived from the variance in measuring the Q's in the two cases, a total of 20 measurements at each frequency. It therefore seems that the effect is real. There is a suggestion of frequency dependence, but hardly a significant one in view of the deviations.

One might ask if the change might be due simply to differences in the viscosity, thermal conductivity, heat capacity, and density of the gas present. These are the parameters that determine wall losses in resonant tubes. Their values for water vapor differ from those for air, but the differences together act in a direction to lower the losses, rather than increase them. When saturated at 24 °C (the average temperature during the tests), water vapor constitutes about 3% of the gas on a molecular basis. Making the assumption that the viscosity, etc., of a mixture can be calculated from the molecular proportions of each of its components, this much water vapor would lower the viscous loss by 0.1% and the thermal loss by 0.4%. The effect therefore is not due to these changes. Molecular absorption attenuation also depends on humidity, but at these frequencies the contribution to loss by molecular absorption is negligible.

The practical significance of this effect in musical instrument theory is probably negligible, at least for the conditions tested here. The effect might, however, be of theoretical interest as a phenomenon that has not, to my knowledge, been addressed before.

[1] N. H.. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer-Verlag, New York, 1991), p. 177.

Predictability of noise indices in a high-rise residential environment (L)

S. K. Tang[a] and W. Y. Chan
Department of Building Services Engineering, The Hong Kong Polytechnic University, Hong Kong, China

(Received 7 November 2002; revised 8 May 2003; accepted 10 June 2003)

In the present investigation, daily noise level measurements were carried out at 12 independent sites within the closely packed high-rise residential areas in Hong Kong. Various environmental noise indices are calculated. Their cross-relationships and correlations with the population density and the daily traffic volume are also examined. Results suggest that the L_{dn} and L_{den} correlate very well with a multiple regression model, which involves both the population density and traffic volume, for the present site conditions. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1596171]

PACS numbers: 43.50.Rq, 43.50.Sr [DKW]

I. INTRODUCTION

Environmental noise has long been an important problem for environmentalists, engineers, and town planners. This problem becomes more acute in densely populated cities with limited availability of land, such as Hong Kong. The population density has also a role in the resultant noise levels. The control of noise exposure, and thus the noise pollution, requires an appropriate noise policy from the local authority, knowledge of the right noise parameters to work with, and also a reliable prediction model.

There have been substantial efforts made in the past few decades on survey of noise from traffic which has been regarded as a very important source of noise in an urbanized city. Empirical noise level prediction models, such as the CRTN,[1] have also been set up. Most of these empirical formulas tend to relate L_{A10} and/or L_{Aeq} with traffic flow volume, vehicle speed, and the percentage of heavy vehicles on the road.

The equivalent sound pressure level and the percentile levels L_{A10} and L_{A90} are important for the estimation of the traffic noise index (TNI),[2] which is believed by many researchers to be an effective predictor of nuisance resulted from noise if they are 24-h averages.[3] However, in the authors' point of view, the 24-h average $L_{A10,T=24\mathrm{hr}}$ will be biased by the noise events that occur during day-time where more than half of the noise sensitive receivers (NSRs) concerned are not at the proximity of the measurement locations. The $L_{A90,T=24\mathrm{hr}}$ depends mostly on night-time events, which may therefore affect more the noise annoyance feeling. A parameter that caters to the different tolerance of the NSRs during a different time of day is therefore required. Parameters such as the day-night level, L_{dn}, and the community noise level, L_{den}, appear to be the legitimate choices.[4] In fact, the use of the former in correlating with annoyance feeling due to noise exposure has been studied extensively by large number of researchers (for instance, Fidell *et al.*[5]).

Despite the established importance of the L_{dn} (and probably the L_{den}) as a descriptor of annoyance caused by environmental noise, the prediction model for it is not thoroughly studied. This study summarizes an attempt to clarify the environmental parameters that are important for the estimation of the L_{dn} in a high-rise urban residential environment where the major sources of noise are the traffic and the population. Discussions on their relationships with the 24-hour averages $L_{A10,T=24\mathrm{hr}}$, $L_{A90,T=24\mathrm{hr}}$ and $L_{Aeq,T=24\mathrm{hr}}$, and the derived levels TNI and the noise pollution level, L_{np},[6] are also presented. A simple multiple regression model will be set up.

II. FORMULATION OF REGRESSION MODELS

The major noise sources in the present study are ground traffic and population. In the Hong Kong urban situation, population and traffic are interrelated, but there may not exist a direct relationship between them. One therefore expects that the noise levels will depend on the traffic volume and average vehicle speed as in most of the traffic noise prediction models. The percentage of heavy vehicles, which has been included in many traffic noise prediction model (for example, CRTN[1]), may not be very relevant to the present study in the residential areas as far as 24-h noise level averages are concerned. The inclusion of this parameter into the present regression model will thus complicate the model without much improvement on the standard error. Also, the recent results from a laboratory study of annoyance caused by traffic noise by Versfeld and Vos[7] suggest the annoyance feeling is independent of the percentage of heavy vehicles.

The regression results of Galloway *et al.*[8] obtained in the United States suggest that the L_{dn} varies with the logarithm of the population density while the effects of traffic have not been explicitly taken into account. Some existing noise level prediction models also include the distance effect. However, the high-rise and closely packed built environment in a congested city like Hong Kong with tall buildings erected on the two sides of a road gives rise to a reverberation effect,[9] which tends to reduce the impact of distance significantly. This site condition is referred to as the ''enclosed'' form by Ko.[10] Therefore and also for the sake of simplicity, the regression model

$$L = a \log_{10} Q + b \log_{10} P + c \quad (\text{model } 1),$$

where L is a 24-h noise index concerned (such as L_{dn} and $L_{Aeq,T=24\mathrm{hr}}$), Q is the daily traffic volume and P is the local

[a]Electronic mail: besktang@polyu.edu.hk, fax: 852-27746146.

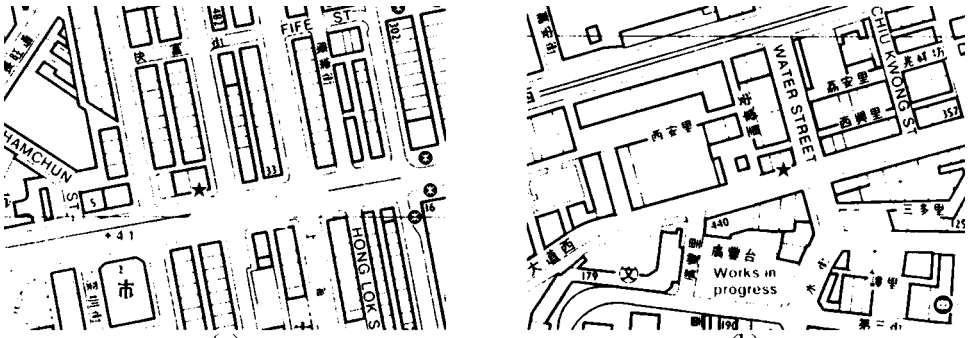


FIG. 1. Typical examples of the selected sites: (a) site D and (b) site E. ⋆: measurement location.

population density, is proposed. The population density in the present study is defined as the number of people living within each square kilometer of area. The coefficient c tends to reflect a prevailing value of L. Correlation of model 1 with various 24-h noise indices will be given and discussed.

III. SITE MEASUREMENTS

In the present site measurement, the sound pressure levels were recorded using the sound level meter Brüel & Kjær 2238F every second at about 1 m from the building façades. Height of measurement varied between low and middle floors. Each measurement lasted for four days, but not on holiday or weekend. A total of 12 sites were surveyed. The sites were all residential ones in the high-rise urban areas or satellite districts of Hong Kong and are basically of the ''enclosed'' form.[10] Two typical site maps are shown in Fig. 1. They are extracted from the official survey maps issued by the local authority and thus some symbols for local uses and some Chinese characters not related to the present study can be found on them.

All the 24-h average noise indices involved in the present study are then calculated from the recorded sound pressure level time series. The daily macroscopic sound pressure level variation pattern at a particular site does not vary very much, so that all the noise indices concerned can be regarded as reasonably stationary (not shown here). The measurements were carried out from November to March during the autumn and winter periods of Hong Kong. Therefore, the noise from domestic air-conditioning is not significant.

The daily average traffic volume data Q and the population density P in the present study were obtained from the official publications of Hong Kong SAR Government.[11,12] It should also be noted that each selected site is closed to only one major road, which is expected to produce most of the traffic noise. Such a road is horizontal or slightly inclined. In addition, there was no screening device between the roads and the sound level measurement points during the measurement period.

IV. RESULTS AND DISCUSSIONS

The major environmental acoustic indices concerned in the present study are the $L_{Aeq,T=24\mathrm{hr}}$, L_{dn}, L_{den}, TNI and L_{np}. The calculation of the first parameter is straightforward. The following formulas are used in the estimation of the others:

$$L_{dn}=10\log_{10}\{\tfrac{1}{24}[16\times 10^{L_{eq(07:00-23:00)}/10}+8\times 10^{(L_{eq(23:00-07:00)}+10)/10}]\}, \quad (1)$$

$$L_{den}=10\log_{10}\{\tfrac{1}{24}[12\times 10^{L_{eq(07:00-19:00)}/10}+4\times 10^{(L_{eq(19:00-23:00)}+5)/10}+8\times 10^{(L_{eq(23:00-07:00)}+10)/10}]\}, \quad (2)$$

$$\mathrm{TNI}=4(L_{A10,T=24\mathrm{hr}}-L_{A90,T=24\mathrm{hr}})+L_{A90,T=24\mathrm{hr}}-30, \quad (3)$$

and

$$L_{np}=L_{Aeq,T=24\mathrm{hr}}+(L_{A10,T=24\mathrm{hr}}-L_{A90,T=24\mathrm{hr}}). \quad (4)$$

TABLE I. Summary of survey data.

Site	Q	P (people/km^2)	Noise indices (dBA)								
			$L_{Aeq,T=24\mathrm{hr}}$	$L_{A10,T=24\mathrm{hr}}$	$L_{A90,T=24\mathrm{hr}}$	L_{dn}	L_{den}	TNI	L_{np}	L_d	L_n
A	37 020	67 171	69.4	73.7	59.0	74.4	74.9	87.8	84.1	70.0	67.9
B	37 580	26 984	70.8	74.6	59.6	74.8	75.3	89.6	85.8	71.7	67.9
C	14 600	80 451	69.0	72.1	60.9	73.5	73.9	75.7	80.2	69.8	66.8
D	31 860	104 305	71.0	75.7	58.1	74.8	75.5	98.5	88.6	72.0	67.6
E	9340	99 508	70.7	74.1	59.3	74.3	74.9	88.5	85.5	71.8	67.0
F	112 750	71 023	73.2	77.3	62.0	76.3	77.0	93.2	88.5	74.5	68.6
G	19 590	38 674	69.1	73.5	59.6	72.6	73.2	85.2	83.0	70.3	65.3
H	10 630	12 793	69.4	73.7	60.5	73.2	73.8	83.3	82.6	70.4	66.1
I	67 670	33 166	71.1	75.3	58.5	73.8	74.5	95.7	87.9	72.4	65.7
J	28 220	926	65.7	70.0	55.9	68.1	69.4	82.3	79.8	66.9	61.0
K	13 160	7411	67.2	69.8	54.3	69.6	70.3	86.3	82.7	68.7	61.0
L	23 680	46 325	69.2	71.5	65.4	73.5	74.0	59.8	75.3	70.1	66.7

TABLE II. Correlations between noise indices.

	$L_{Aeq,T=24\text{hr}}$	$L_{A10,T=24\text{hr}}$	$L_{A90,T=24\text{hr}}$	L_{dn}	L_{den}	TNI	L_{np}	L_d	L_n
$L_{Aeq,T=24\text{hr}}$	1								
$L_{A10,T=24\text{hr}}$	0.944	1							
$L_{A90,T=24\text{hr}}$	0.476	0.325	1						
L_{dn}	0.936	0.869	0.616	1					
L_{den}	0.950	0.897	0.600	0.997	1				
TNI	0.443	0.618	0.542	0.298	0.298	1			
L_{np}	0.689	0.800	0.292	0.516	0.550	0.954	1		
L_d	0.992	0.934	0.418	0.886	0.902	0.483	0.721	1	
L_n	0.843	0.793	0.680	0.973	0.967	0.139	0.382	0.769	1

Two additional noise levels, namely the day level L_d and night level L_n, are also calculated in the present study as supplementary data and are discussed whenever necessary. Their definitions are

$$L_d = L_{eq(07:00-23:00)} \quad \text{and} \quad L_n = L_{eq(23:00-07:00)}. \tag{5}$$

The definitions of day and night here follow the local legislation.

Table I summarizes the data obtained in the present study. Site J is a new high-rise town and the population there at the time of measurement had not yet reached its expected steady-state value. One can verify immediately that there is no correlation between $\log_{10} Q$ and $\log_{10} P$ ($R=0.12$, $p=0.7$). The magnitudes of the linear correlation coefficients, R, between any two listed indices are presented in Table II. It can clearly be observed that the correlation between the L_{Aeq} and L_{dn} (and also L_{den}) is very high. Since the L_{dn} and L_{den} have good correlations with annoyance due to noise exposure,[5] the present correlation result appears to be consistent with the findings of Versfeld and Vos[7] which show that the L_{Aeq} is a good indicator for noise annoyance. The correlations between TNI and other indices are in general very weak.

The correlations between the present listed noise indices and $\log_{10} Q$ are in general not strong. The higher correlation coefficients are those with the L_d and $L_{A10,T=24\text{hr}}$, which are only around 0.3 (Table III). This indicates that the ground transportation does not dominate the overall acoustical environment of the residential areas. However, the correlation data does suggest Q is slightly more important in affecting the noise level during daytime (07:00 to 23:00). The corresponding correlations with $\log_{10} P$ are more impressive as shown in Table III. The relatively higher correlations with P are probably due to the noise sources related to the mixed modes of businesses being carried out in the residential areas. These businesses include shops (usually at ground levels), markets, schools, and restaurants. Sometimes, one can find shopping malls of various scales in these residential areas. Reverberation in the street canyons is also one of the reasons. The higher the population density, the closer the packing of buildings expected. Speeds of vehicles in these areas are under stringent control. Certainly, further investigation will be helpful on this aspect.

Table IV summarizes the results of the present multiple regressions at 95% confidence level. The corresponding statistical F-test parameter for the multiple regression $F_{2,9,0.95}$ is 3.01 and the numbers in parentheses refer to the standard errors of the respective regression coefficients. The coefficients with noise indices having $p<0.05$ in general survive the statistical t-test (not shown here) and their percentage errors are usually relatively small. One can find immediately that, despite the possibility of variations in site conditions, the correlations related to $L_{Aeq,T=24\text{hr}}$, L_{dn}, and L_{den} are very high. One can also notice from Tables III and IV that the corresponding correlations are improved when both Q and P are taken into account together, suggesting that the contributions from Q and P in the overall acoustical environment are competitive. The inclusion of nominal vehicle speed V into the regression model 1 deteriorates considerably the correlations (not shown here). This may be due to the small range of $\log_{10} V$ in the present study ($50\,\text{km/h} \leqslant V \leqslant 80\,\text{km/h}$) and/or the effects of V have been already catered for by Q and/or P under the long term noise level averaging. The partial coefficients of correlations suggest that the population density P is relatively more dominating in affecting the energy based noise indices, such as $L_{Aeq,T=24\text{hr}}$ and L_{dn}.

V. CONCLUSIONS

Twenty-four-hour noise level measurements were carried out at 12 independent sites within the closely packed high-rise residential areas of Hong Kong, where the major sources of noise were the ground traffic and the population. Each measurement lasted for four consecutive days. Various noise indices are calculated and their correlations with the population densities P and the daily traffic volume Q are examined. Sites with noise screening devices were excluded from the present study.

It is found in the present survey that the population density is in general better correlated with the energy based

TABLE III. Correlations between noise indices with site parameters Q and P.

	$L_{Aeq,T=24\text{hr}}$	$L_{A10,T=24\text{hr}}$	$L_{A90,T=24\text{hr}}$	L_{dn}	L_{den}	TNI	L_{np}	L_d	L_n
$\log_{10} Q$	0.519	0.575	0.149	0.393	0.437	0.387	0.476	0.531	0.329
$\log_{10} P$	0.763	0.647	0.535	0.874	0.850	0.140	0.362	0.712	0.850

TABLE IV. Multiple regression results (model 1: $L = a \log_{10} Q + b \log_{10} P + c$).

Noise index	Regression formula coefficients[a]			Partial correlation coefficient		Standard error (dB)	Statistical test parameters		
	a	b	c	$r_{L\cdot\log_{10} Q\|\log_{10} P}$	$r_{L\cdot\log_{10} P\|\log_{10} Q}$		R	F	P
$L_{Aeq,T=24\text{hr}}$	2.59	2.33	47.7	0.428	0.704	0.6	0.874	14.62	0.0015
	(0.98)	(0.54)	(4.7)						
$L_{A10,T=24\text{hr}}$	3.51	2.23	48.0	0.499	0.581	0.8	0.818	9.07	0.0070
	(1.35)	(0.74)	(6.5)						
$L_{A90,T=24\text{hr}}$	0.74	2.50	45.0	0.084	0.520	1.5	0.514	1.87	0.2098
	(11.75)	(1.34)	(11.8)						
L_{dn}	2.04	3.23	49.8	0.287	0.832	0.6	0.920	24.90	0.0002
	(0.93)	(0.51)	(4.4)						
L_{den}	2.22	2.92	51.0	0.334	0.802	0.6	0.914	22.74	0.0003
	(0.90)	(0.49)	(4.3)						
TNI	11.83	1.44	26.9	0.374	0.083	5.9	0.396	0.837	0.4641
	(9.68)	(5.30)	(46.3)						
L_{np}	5.37	2.07	50.7	0.434	0.306	2.1	0.566	2.116	0.1765
	(1.86)	(3.40)	(16.3)						
L_d	2.71	2.16	49.1	0.447	0.651	0.7	0.840	10.82	0.0040
	(0.60)	(1.09)	(5.2)						
L_n	1.77	3.49	42.6	0.226	0.815	0.8	0.879	15.35	0.0013
	(0.68)	(1.24)	(5.9)						

[a]Numbers in parentheses are the standard errors of the regression coefficients.

noise indices like the day level, night level, $L_{Aeq,T=24\text{hr}}$, L_{dn}, and L_{den}, especially the latter two, which have been proposed to be highly related to the human annoyance by other researchers. An important observation in the present study is the substantial improvement in the correlations between the noise indices L_{dn} and L_{den} and the environmental parameters Q and P when both Q and P are taken into account by a simple multiple regression model. The inclusion of vehicle speed results in weaker correlations. It is also found that the energy-based noise indices are more sensitive to a change in the population density.

ACKNOWLEDGMENTS

This project is supported by a grant from the Research Grant Council, Hong Kong SAR Government. (Project No. PolyU 5024/02E.) The assistance of H. Y. Poon and C. H. Ng in the data collection is very much appreciated.

[1] UK Department of Transport, *Calculation of Road Traffic Noise* (Her Majesty's Stationary Office, London, 1988).

[2] W. E. Scholes, "Traffic noise criteria," Appl. Acoust. **3**, 1–21 (1970).

[3] I. D. Griffiths and F. J. Langdon, "Subjective response to road traffic noise," J. Sound Vib. **8**, 16–32 (1968).

[4] T. J. Schultz, *Community Noise Ratings* (Applied Science, London, 1972).

[5] S. Fidell, D. S. Barber, and T. J. Schultz, "Updating a dosage-effect relationship for the prevalence of annoyance due to general transportation noise," J. Acoust. Soc. Am. **89**, 221–233 (1991).

[6] D. W. Robinson, "Towards a unified system of noise assessment," J. Sound Vib. **14**, 279–298 (1971).

[7] N. J. Versfeld and J. Vos, "*A*-weighted equivalent sound levels as a predictor of the annoyance caused by road traffic consisting of various proportions of light and heavy vehicles," J. Sound Vib. **253**, 389–399 (2002).

[8] W. J. Galloway, K. M. Elderd, and M. A. Simpson, "Population distribution of the United States as a function of outdoor noise level," EPA Report 550/9-74-900 (1974).

[9] N. W. M. Ko and C. P. Tang, "Reverberation time in a high-rise city," J. Sound Vib. **56**, 459–461 (1978).

[10] N. W. M. Ko, "Traffic noise in a high-rise city," Appl. Acoust. **11**, 225–239 (1978).

[11] Transport Department, *The Annual Traffic Census 2001* (Hong Kong SAR Government, 2002).

[12] Census and Statistics Department, *2001 Population Census: Summary Results* (Hong Kong SAR Government, 2001).

Ray dynamics in a long-range acoustic propagation experiment

Francisco J. Beron-Vera and Michael G. Brown
Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida 33149

John A. Colosi
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Steven Tomsovic
Department of Physics, Washington State University, Pullman, Washington 99164

Anatoly L. Virovlyansky
Institute of Applied Physics, Russian Academy of Science, 6003600 Nizhny Novgorod, Russia

Michael A. Wolfson
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

George M. Zaslavsky
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012

(Received 5 September 2002; revised 8 May 2003; accepted 23 June 2003)

A ray-based wave-field description is employed in the interpretation of broadband basin-scale acoustic propagation measurements obtained during the Acoustic Thermometry of Ocean Climate program's 1994 Acoustic Engineering Test. Acoustic observables of interest are wavefront time spread, probability density function (PDF) of intensity, vertical extension of acoustic energy in the reception finale, and the transition region between temporally resolved and unresolved wavefronts. Ray-based numerical simulation results that include both mesoscale and internal-wave-induced sound-speed perturbations are shown to be consistent with measurements of all the aforementioned observables, even though the underlying ray trajectories are predominantly chaotic, that is, exponentially sensitive to initial and environmental conditions. Much of the analysis exploits results that relate to the subject of ray chaos; these results follow from the Hamiltonian structure of the ray equations. Further, it is shown that the collection of the many eigenrays that form one of the resolved arrivals is nonlocal, both spatially and as a function of launch angle, which places severe restrictions on theories that are based on a perturbation expansion about a background ray. [DOI: 10.1121/1.1600724]

PACS numbers: 43.20.Dk, 43.30.Cq [RAS]

I. INTRODUCTION

The chaotic dynamics of ray trajectories in ocean acoustics have been explored in a number of recent publications,[1–22] including two recent reviews.[21,22] In these publications a variety of theoretical results are presented and illustrated, mostly using idealized models of the ocean sound channel. In contrast, in the present study a ray-based wave-field description is employed, in conjunction with a realistic environmental model, to interpret a set of underwater acoustic measurements. For this purpose we use both oceanographic and acoustic measurements collected during the November 1994 Acoustic Engineering Test (the AET experiment). In this experiment phase-coded pulse-like signals with 75-Hz center frequency and 37.5-Hz bandwidth were transmitted near the sound-channel axis in the eastern North Pacific Ocean. The resulting acoustic signals were recorded on a moored vertical receiving array at a range of 3252 km. We present evidence that all of the important features of the measured AET wave fields are consistent with a ray-based wave-field description in which ray trajectories are predominantly chaotic.

The AET measurements have previously been analyzed in Refs. 23–25. In the analysis of Worcester *et al.*[23] it was shown that the early AET arrivals could be temporally resolved, were stable over the duration of the experiment, and could be identified with specific ray paths. These features of the AET observations are consistent with those of other long-range data sets.[26,27] The AET arrivals in the last 1 to 2 s (the arrival finale) could not be temporally resolved or identified with specific ray paths, however. Also, significant vertical scattering of acoustic energy was observed in the arrival finale. These features of the AET arrival finale are consistent with measurements and simulations (both ray- and parabolic-equation-based) at 250-Hz and 1000-km range[27–29] and at 133-Hz and 3700-km range.[30]

The analysis of Colosi *et al.*[24] focused on statistical properties of the early resolved AET arrivals. Measurements of time spread (a measure of scattering-induced pulse broadening), travel time variance (a measure of scattering-induced pulse wander), and probability density functions (PDFs) of peak intensity were presented and compared to theoretical predictions based on a path integral formulation as described in Ref. 31. Travel time variance was well predicted by the theory, but time spreads were two to three orders of magnitude smaller than theoretical predictions, and peak intensity PDFs were close to lognormal, in marked contrast to the

predicted exponential PDF. The surprising conclusion of this analysis was that the measured AET pulse statistics suggested propagation in or near the unsaturated regime (characterized physically by the absence of micromultipaths and mathematically by use of a perturbation analysis based on the Rytov approximation), while the theory predicted propagation in the saturated regime (characterized by the presence of a large number of micromultipaths whose phases are random).

In Ref. 25 it was shown that in the finale region of the measured AET wave fields, where no timefronts are resolved, the intensity PDF is close to the fully saturated exponential distribution. This result is not unexpected because the finale region is characterized by strong scattering of both rays[14] and modes.[32]

The work reported here seeks to elucidate both the physics underlying the AET measurements, and the causes of the successes and failures of the analyses reported in Refs. 23–25. We show that a ray-based analysis, in which ideas associated with ray chaos play an important role, can account for the stability of the early arrivals, their small time spreads, the associated near-lognormal PDF for intensity peaks, the vertical scattering of acoustic energy in the reception finale, and the near-exponential intensity PDF in the reception finale. Part of the success of this interpretation comes about due to the presence of a mixture of stable and unstable ray trajectories. Also, an explanation is given for differing intensity statistics in the early and late portions of the measured wave fields, and the fairly rapid transition between these regimes.

The remainder of this paper is organized as follows. In the next section we describe the AET environment and the most important features of the measured and simulated wave fields. In the simulated wave fields both measured mesoscale and simulated internal-wave-induced sound-speed perturbations are taken into account. It is shown that even in the absence of internal waves, the late arrivals are not resolved and the associated ray paths are chaotic. In Sec. III we present results that relate to the structure of the early portion of the measured timefront and its stability. The micromultipathing process is discussed, and a quantitative explanation for the remarkably small time spreads of the early ray arrivals is given. Section IV is concerned with intensity statistics. An explanation is given for the cause of the different wave-field intensity statistics in the early and late portions of the arrival pattern. In the final section we summarize and discuss our results.

II. MEASURED AND SIMULATED WAVE FIELDS

Figure 1 shows a measured AET wave field in the time–depth plane and ray-based simulations of such a wave field with and without internal waves. Figure 2 is a complementary display that shows measured and simulated wave fields after plane-wave beamforming. In this section we describe the most important qualitative features of measured and simulated AET wave fields. We begin by reviewing fundamental ray theoretical results so that our ray simulations can be fully understood. Also, this material provides a foundation for some of the later discussion. We then briefly describe the AET environment and the signal-processing steps that were

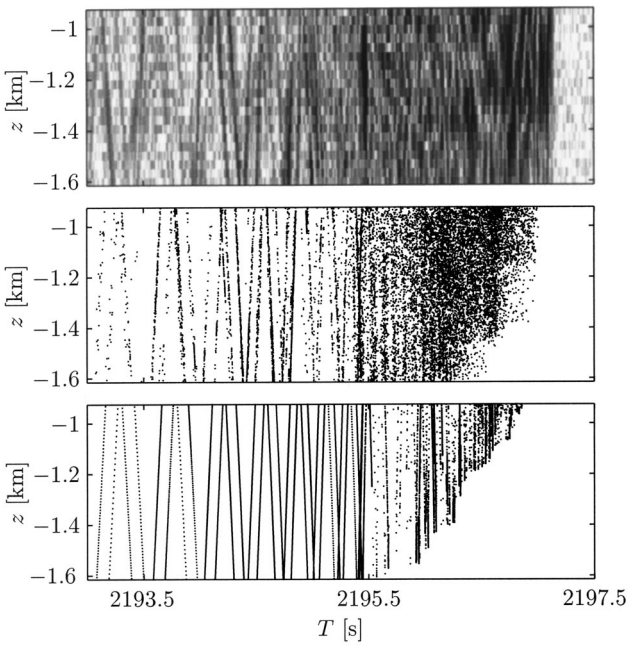


FIG. 1. Measured and simulated AET wave fields in the time–depth plane. Upper panel: a typical measurement, shown on a gray-scale plot with a dynamic range of 30 dB, of wave-field intensity. Middle panel: ray simulation with internal waves. Lower panel: ray simulation without internal waves. Wave-field intensity in the simulations is approximately proportional to the local density of dots, each corresponding to a ray.

performed to produce the measured wave field shown in Fig. 1. As is common practice in ocean acoustics, we treat mesoscale ocean structure [whose horizontal length scales are O(100 km) and whose time scales are O(1 month)] determin-

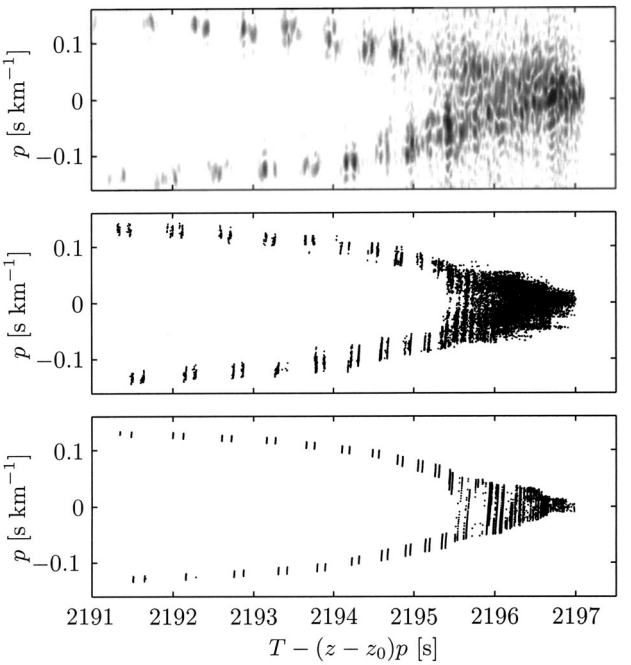


FIG. 2. Same as Fig. 1 except that plane-wave beamformed wave fields are shown. The beamformed ray simulations assume a dense receiving array whose upper and lower bounds coincide with the bounds of the AET receiving array. The reference depth was taken to be the depth of the uppermost hydrophone, $z_0 = -0.9$ km. The vertical slowness p is defined in the text following Eq. (7).

istically, and internal gravity wave variability [whose horizontal length scales are O(1 km) or more and whose time scales are O(10 min) to O(1 day)] using a statistical description. Our treatment of internal-wave-induced sound-speed perturbations is briefly described. Finally, we return to describing and comparing qualitative features of the measured and simulated wave fields.

A. Theoretical background

Fixed-frequency (cw) acoustic wave fields $u(z,r;\sigma)$ satisfy the Helmholtz equation

$$\nabla^2 u + \sigma^2 c^{-2}(z,r)u = 0, \tag{1}$$

where $\sigma = 2\pi f$ is the angular frequency of the wave field and $c(z,r)$ is the sound speed. (The extension to transient wave fields is straightforward using Fourier synthesis. It is not necessary to consider this complication to present the most important results needed below.) We assume propagation in the vertical plane with Cartesian coordinates z, increasing upwards with $z=0$ at the sea surface, and range r. Ray methods (see, e.g., Ref. 33 for a thorough discussion in a broader context) can be introduced when the wavelength $2\pi c/\sigma$ of all waves of interest is small compared to the length scales that characterize variations in c. Substituting the geometric ansatz

$$u(z,r;\sigma) = \sum_j A_j(z,r) e^{i\sigma T_j(z,r)}, \tag{2}$$

representing a sum of locally plane waves, into the Helmholtz equation gives, after collecting terms in descending powers of σ, the eikonal equation

$$(\nabla T)^2 = c^{-2}, \tag{3}$$

and the transport equation

$$\nabla(A^2 \nabla T) = 0. \tag{4}$$

For notational simplicity, the subscripts on A and T have been dropped in (3) and (4). Each term in the sum (2) represents an ''eigenray'' connecting a source and receiver, as discussed below. More generally, each A_j in (2) can be thought of as the leading term in an asymptotic expansion in powers of σ^{-1}, but the higher-order terms in such an expansion play no role in our analysis.

The solution to the eikonal equation (3) can be reduced to solving a set of ray equations. If a one-way formulation, valid when r increases monotonically following all rays of interest (see, e.g., Ref. 22), is invoked, the ray equations are

$$\frac{dz}{dr} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dr} = -\frac{\partial H}{\partial z}, \tag{5}$$

and

$$\frac{dT}{dr} = L = p\frac{dz}{dr} - H, \tag{6}$$

where

$$H(p,z,r) = -\sqrt{c^{-2}(z,r) - p^2}. \tag{7}$$

These equations constitute a nonautonomous Hamiltonian system with one degree of freedom; z and p are canonically conjugate position and momentum variables, r is the independent variable, H is the Hamiltonian, L is the Lagrangian, and the travel time T corresponds to the classical action. The system is nonautonomous because H depends explicitly on the independent variable r. Note also that $p \equiv p_z = \partial T/\partial z$ and $p_r = \partial T/\partial r = -H$ are the z and r components, respectively, of the ray slowness vector. The ray angle relative to the horizontal φ satisfies $dz/dr = \tan\varphi$, or, equivalently, $cp = \sin\varphi$. For a point source, rays can be labeled by their initial slowness p_0. The solution to the ray equations is then $z(r;p_0)$, $p(r;p_0)$, $T(r;p_0)$. The terms in (2) correspond to eigenrays, which are solutions to the two-point boundary value problem $z(r_R;p_0) = z_R$, where the spatial coordinates of the receiver are (z_R, r_R). The transport equation (4) can be reduced to a statement of constancy of energy flux in ray tubes. The solution for the jth eigenray can be written

$$A_j(z,r) = A_0 |H q_{21}|_j^{-1/2} e^{-i\mu_j(\pi/2)}, \tag{8}$$

where H, the matrix element q_{21}, and the Maslov index μ are evaluated at the endpoint of the ray, and A_0 is a constant. The stability matrix

$$Q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} \left.\frac{\partial p}{\partial p_0}\right|_{z_0} & \left.\frac{\partial p}{\partial z_0}\right|_{p_0} \\ \left.\frac{\partial z}{\partial p_0}\right|_{z_0} & \left.\frac{\partial z}{\partial z_0}\right|_{p_0} \end{pmatrix}, \tag{9}$$

quantifies ray spreading

$$\begin{pmatrix} \delta p \\ \delta z \end{pmatrix} = Q \begin{pmatrix} \delta p_0 \\ \delta z_0 \end{pmatrix}. \tag{10}$$

Elements of this matrix evolve according to

$$\frac{d}{dr} Q = KQ, \tag{11}$$

where Q at $r=0$ is the identity matrix, and

$$K = \begin{pmatrix} -\frac{\partial^2 H}{\partial z \partial p} & -\frac{\partial^2 H}{\partial z^2} \\ \frac{\partial^2 H}{\partial p^2} & \frac{\partial^2 H}{\partial z \partial p} \end{pmatrix}. \tag{12}$$

At caustics q_{21} vanishes and, for waves propagating in two space dimensions, the Maslov index advances by one unit.

In ocean environments with realistic range dependence, including the AET environment, ray trajectories are predominantly chaotic. Chaotic rays diverge from neighboring rays at an exponential rate, on average, and are characterized by a positive Lyapunov exponent

$$\nu_L = \lim_{r\to\infty} \left(\frac{1}{r} \lim_{\mathcal{D}(0)\to 0} \ln \frac{\mathcal{D}(r)}{\mathcal{D}(0)} \right). \tag{13}$$

Here, $\mathcal{D}(r)$ is a measure of the separation between rays at range r; suitable choices for $\mathcal{D}$ are any of the four matrix elements of Q or the trace of Q. The chaotic motion of ray

trajectories leads to some limitations on predictability. This will be discussed in more detail below. Additional background material relating to this topic can be found in Refs. 21, 22 and references therein.

B. The AET experiment

In the AET experiment a source, suspended at 652 m depth from the floating instrument platform R/P FLIP in deep water west of San Diego, transmitted sequences of a phase-coded signal whose autocorrelation was pulse-like (with a pulse width of approximately 27 ms). The source center frequency was 75 Hz and the bandwidth was approximately 37.5 Hz. The resulting acoustic signals were recorded east of Hawaii at a range of 3252.38 km on a moored 20-element receiving array between the depths of 900 and 1600 m. After correcting for the motion of the source and receiving array, including removal of associated Doppler shifts, the received acoustic signals were crosscorrelated with a replica of the transmitted signal (to achieve the desired pulse compression) and complex demodulated. To improve the signal-to-noise ratio, 28 consecutive processed receptions, extending over a total duration of 12.7 min, were coherently added. The duration over which this coherent processing yields improved signal to noise is limited by ocean fluctuations, mostly due to internal waves. The statistics of the acoustic fluctuations were shown in Ref. 25 not to be adversely affected by the 12.7-min coherent integration. Nearly concurrent with the week-long experiment, temperature profiles in the upper 700 m were measured with XBTs every 30 km along the transmission path. An objective mapping technique was applied[23] to these measurements to produce a map of the background sound-speed structure (including mesoscale structure) along the transmission path. The sampling interval in range of the resulting sound-speed field is 30 km so that no structure with horizontal wavelengths less than 60 km is present in this field. More details on the AET experiment, environment, and processing of the acoustic signals can be found in Ref. 23.

Internal-wave-induced sound-speed perturbations in our simulations are assumed to satisfy the relationship $\delta c = (\partial c/\partial z)_p \zeta$, and the statistics of ζ, the internal-wave-induced vertical displacement of a water parcel, are assumed to be described by the empirical Garrett–Munk spectrum (see, e.g., Ref. 34). The potential sound-speed gradient $(\partial c/\partial z)_p$, the buoyancy frequency N, and the sound speed c were estimated directly from hydrographic measurements using a procedure similar to that described in the Appendix of Ref. 30; it was found that a good approximation for the AET environment is $(\partial c/\partial z)_p/c \approx (1.25\,\mathrm{s^2/m})N^2$. Individual realizations of the vertical displacement $\zeta(r,z)$ were computed using Eq. (19) of Ref. 35 with the variable x used to denote the range variable, and $y=t=0$. Physically, this corresponds to a frozen vertical slice of the internal wave field that includes the influence of transversely propagating internal wave modes. A Fourier method was used to numerically generate sound-speed perturbation fields within a 32 768-point, 3276.8-km domain in range—so $\Delta x = 0.1$ km and $\Delta k_x = 2\pi/3276.8$ km. A hard internal wave horizontal wave number cutoff, $|k_x| \leqslant k_{\max} = 2\pi/1$ km, was used in all of the simulations shown. (Ray calculations were also performed using internal wave fields in which the hard upper bound on k_x was replaced by an exponential damping of the large k_x energy. The qualitative ray behavior described below did not depend on the manner in which the large k_x cutoff was treated.)

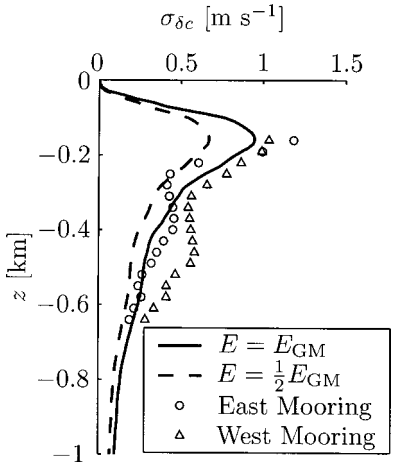


FIG. 3. Measured and simulated rms sound-speed fluctuations in the AET environment. The measurements shown were made approximately 1000 km northwest of the source (labeled east) and approximately 1000 km northeast of the receiving array (labeled west). Simulated rms fluctuations due to internal waves are shown for both $E = 0.5\,E_{\mathrm{GM}}$ and $E = 1.0\,E_{\mathrm{GM}}$, as described in the text.

A mode number cutoff of $j_{\max} = 30$ was used in all the simulations shown. Contributions from approximately 2^{30} ($2^{15}/5 \times 2^{15}/5 \times 30$, where the factors of 1/5 are present because the assumed wave number cutoff was one-fifth of the assumed Nyquist wave number) internal wave modes were included in our simulated internal-wave-induced sound-speed perturbation fields. A sparse set of measurements of sound-speed variance made near the midpoint of the transmission path during the AET experiment[24] suggests that the internal wave energy strength parameter E was close to half the nominal Garrett–Munk value E_{GM}. More extensive and accurate measurements (see Fig. 3) made in 1998/1999 at two locations along a path displaced about 800 km to the north of the AET transmission path favor a value of E close to E_{GM}. Ray simulations were performed using both $E = 0.5\,E_{\mathrm{GM}}$ and $E = 1.0\,E_{\mathrm{GM}}$. All of the simulations shown in this paper use $E = 0.5\,E_{\mathrm{GM}}$. The difference between these simulations and those performed using $E = 1.0\,E_{\mathrm{GM}}$ (not shown) will be discussed where appropriate.

C. Simulated wave fields

With the foregoing discussion as background, we are prepared to discuss our ray simulations in the AET environment. These are shown in Figs. 1, 2, 4–11, and 14. The points plotted in Figs. 1, 2, and 4 were computed by integrating the coupled ray equations (5) and (6) from the source to the range of the receiving array. Several fixed and variable step-size integration algorithms were tested. In the presence of internal waves, convergence using a fixed-step fourth-order Runge–Kutta integrator required that the step size not exceed approximately 1 meter. All ray-tracing calculations presented in this paper were performed using double-

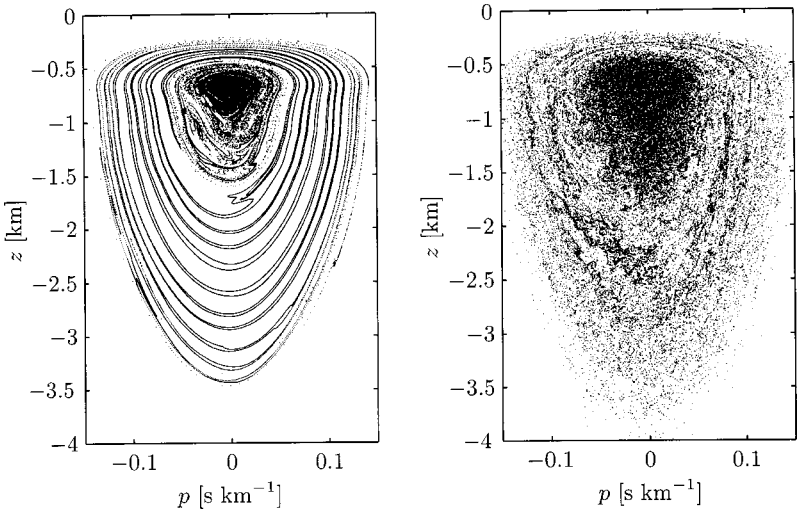


FIG. 4. Simulated Lagrangian manifolds in the AET environment without (left panel) and with (right panel) a superimposed internal-wave-induced sound-speed perturbation field. The ray density used in the plot on the right and for small $|p|$ in the plot on the left is too sparse to resolve what should, in each plot, be a smooth unbroken curve that does not intersect itself.

precision arithmetic (64-bit floating point wordsize). The wave-field intensity is approximately proportional to the density of rays (dots) that are plotted in Figs. 1 and 2. A somewhat more difficult calculation in which the contributions to the wave field from many rays are coherently added will be discussed below.

In the presence of internal waves, rays with launch angles between approximately $\pm 5°$ form the diffuse finale region of the arrival pattern arriving after approximately 2196 s. The sound speed at the source was approximately 1.48 km/s, so $|\varphi_0| \leqslant 5°$ corresponds approximately to $|p_0| \leqslant 0.06$ s/km. Steeper rays contribute to the earlier, mostly resolved arrivals. It is seen in Fig. 1 that our ray simulation in the presence of internal waves slightly underestimates the vertical spread of the finale region, although an ensemble average over several pulses would be needed to quantify the effect. Our experience[29] suggests that the variability of the spread is of the order of the discrepancy seen in Fig. 1, and that some details of the internal wave field may be important in controlling this spread. We note also that a slight underestimate in our ray simulation is expected inasmuch as diffractive effects that are not included in this simulation contribute approximately an additional 200-m vertical spread of energy at 75 Hz. (This number can be estimated from parabolic equation simulations, examination of lower-order modes, or a local Airy function analysis.) With these caveats, the agreement between simulated and measured vertical spreads of energy in the finale region is fairly good. A striking feature of Figs. 1, 2, and 4 is the contrast between the diffuse distribution of steep rays (with large p and z excursions) in (p,z) seen in Fig. 4 when internal waves are present and the tightly clustered distributions of the same rays in the time–depth (Fig. 1) and time–angle (Fig. 2) plots.

The cause of the complexity seen in Fig. 4 is ray chaos;[21,22] most ray trajectories diverge from initially infinitestimally perturbed rays at an exponential rate. A total of 48 000 rays, sampled uniformly in launch angle between $\pm 12°$, was traced to produce Figs. 1, 2, and 4 in the presence of internal waves. This fan of rays is far too sparse to resolve what should be an unbroken smooth curve—a Lagrangian manifold—which does not intersect itself in phase space (Fig. 4). Under chaotic conditions the separation between neighboring rays grows exponentially, on average, in range. The complexity of the Lagrangian manifold grows at the same exponential rate. The Lyapunov exponent is the reciprocal of the e-folding distance (see, e.g., Ref. 22). Finite range numerical estimates of Lyapunov exponents (hereafter referred to as stability exponents) are shown as a function of launch angle in Fig. 5. It is seen that in this environment the flat rays ($|\varphi_0| \lesssim 5°$) have stability exponents of approximately 1/(100 km), while the steeper rays ($6° \lesssim |\varphi_0| \lesssim 11°$) have stability exponents of approximately 1/(300 km). It follows, for example, that the complexity of the flat ray portion of the Lagrangian manifold grows approximately like $\exp(r/100\ \text{km})$.

Not surprisingly, stability exponents show some sensitivity to parameters that describe the internal wave field. Our simulations show that the stability exponents of rays with axial angles of approximately 5° or less approximately double when $j_{\max}$ increases from 30 to 100, while steeper rays show very little sensitivity to $j_{\max}$. In contrast, stability exponents of rays with axial angles from approximately 8° to 15° approximately double when $k_{\max}$ increases from $2\pi/1000$ m to $2\pi/250$ m, while the flatter rays show little sensitivity to $k_{\max}$. With these comments in mind, one should not attach too much significance to the details of Fig. 5. We note, however, that in all of the simulations performed, rays with axial angles of approximately 5° or less had higher stability exponents than rays in the 5° to 10° band.

Note that what appears to be temporally resolved arriv-

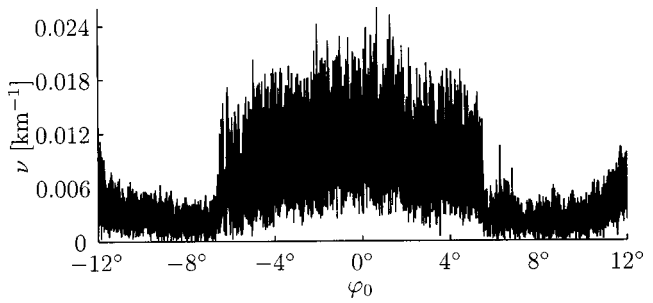


FIG. 5. Stability exponents (finite range estimates of Lyapunov exponents) as a function of ray launch angle in the AET environment including a superimposed internal-wave-induced sound-speed perturbation field.

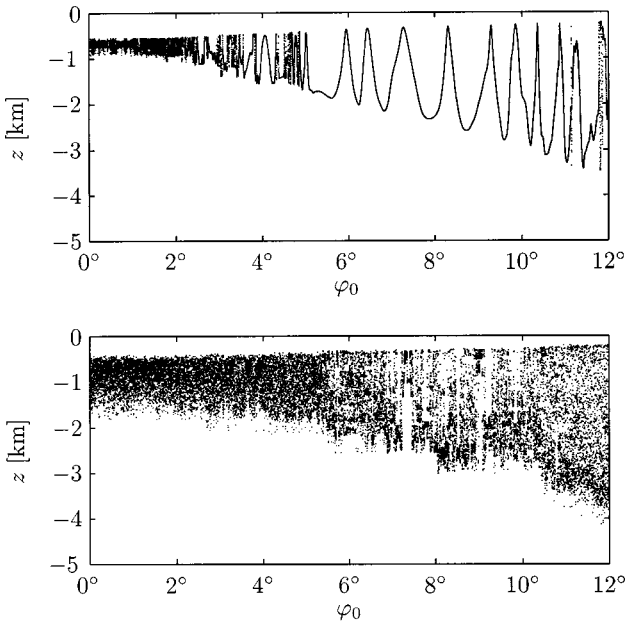


FIG. 6. Ray depth vs launch angle in the AET environment without (upper panel) and with (lower panel) a superimposed internal-wave-induced sound-speed perturbation field. In both panels $\Delta\varphi_0 = 0.0005°$.

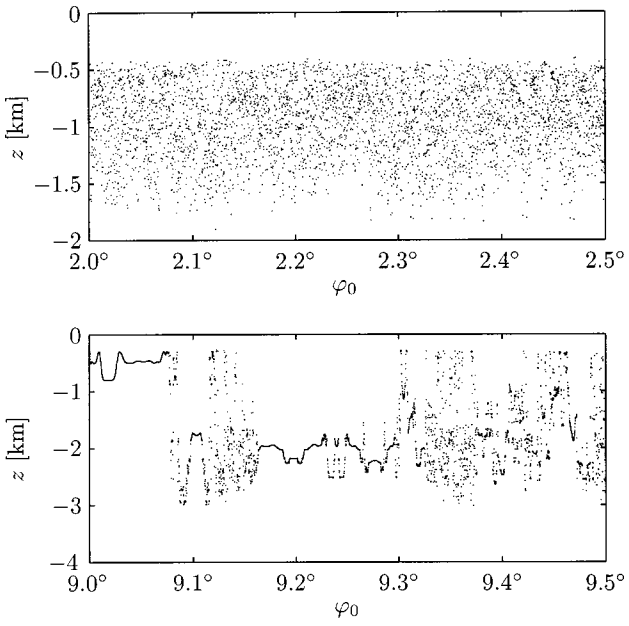


FIG. 7. Expanded view of the lower panel of Fig. 6 showing ray depth vs launch angle in two small angular bands in the AET environment including a superimposed internal-wave-induced sound-speed perturbation field. In both panels $\Delta\varphi_0 = 0.000\,125°$.

als in the measured wave field correspond in our simulations to contributions from an exponentially large number of ray paths. (Only a small fraction of the total number is included in our simulations, however, because of the relative sparseness of our initial set of rays.) The observation that the travel times of chaotic ray paths may cluster and be relatively stable was first made in Ref. 8.

Simmen *et al.*[14] have previously produced plots similar to our Figs. 1 and 4 for ray motion in a deep ocean model consisting of a background sound-speed structure very similar to ours on which internal-wave-induced sound-speed perturbations were superimposed. Although our results are similar in many respects, it is noteworthy that the right panel of our Fig. 4 shows more chaotic behavior than is present in the corresponding plot in Ref. 14. This difference persists if our 3252.38-km range simulations are replaced by simulations at the same range (1000 km) that was used in Ref. 14. Our rays, especially the steeper rays, are more chaotic than those in Ref. 14. This difference is likely due to the increased complexity of our internal wave field over that of Simmen *et al.*,[14] who included contributions from ten internal wave modes in their simulated internal wave fields.

Figures 6 and 7 show plots of ray depth at the AET range vs launch angle. In Fig. 6 the same rays (excluding those with negative launch angles) that were used to produce Figs. 1, 2, and 4 are plotted. In Fig. 7 two small angular bands of the upper panel of Fig. 6 are blown up; the ray density is four times greater than that used in Fig. 6. Figure 6 strongly suggests that, even in the absence of internal waves, the near-axial rays are chaotic. Not surprisingly, in the presence of internal waves, steeper rays are also predominantly chaotic. Note, however, that Figs. 5, 6, and 7 show that the near-axial rays are evidently more chaotic than the steeper rays. An explanation for this behavior that relates to the background sound-speed structures will be given in Sec. IV.

In this subsection we have briefly described and compared the gross features of measured and simulated AET wave fields. Ray trajectories in our simulated wave fields in the presence of internal waves are predominantly chaotic. In spite of this, Figs. 1 and 2 show that many features of our simulated wave fields are stable and in good qualitative agreement with the observations. Simulated and measured spreads of acoustic energy in time, depth, and angle are generally in good agreement, both in the early and late portions of the arrival pattern. We shall not discuss further the spreads of energy in depth and angle, except to note that Figs. 1 and 2 show that these spreads can be accounted for using ray methods in the presence of realistic (including internal-wave-induced sound-speed perturbations) ocean structure. In the two sections that follow, we shall consider in more detail the time spreads of the early ray arrivals, and intensity statistics in both the early and late portions of the arrival pattern.

III. MICROMULTIPATHS AND TIMEFRONT STABILITY

In this section we consider in some detail eigenrays, timefront stability, and time spreads. We focus our attention on the early branches of the timefront, where measured time spreads were, surprisingly, only approximately 2 ms, in sharp contrast to the theoretical prediction of approximately 1 s.[24] We focus our attention on the early arrival time spreads because measured time-spread estimates are available only for these arrivals. In the first subsection we discuss micromultipaths, and describe the important qualitative features of the micromultipathing process. Also, we explain how the micromultipath travel time spreads of O(10 ms) in our simulations can be reconciled with O(1 ms) measured wave-field time spreads. In the second subsection a quantitative estimate of

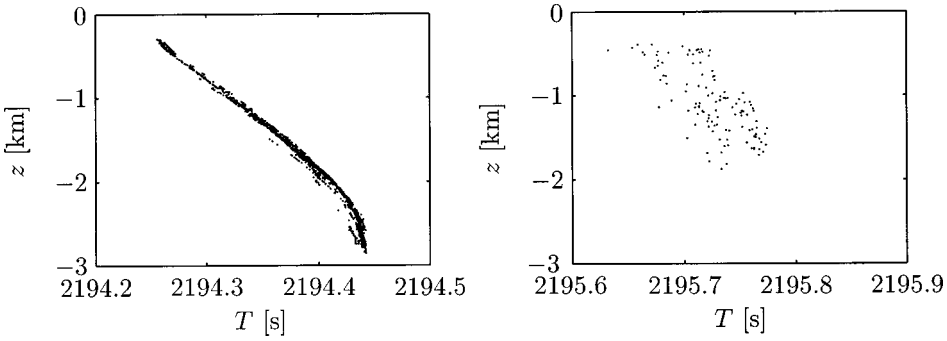


FIG. 8. Ray depth vs travel time in the AET environment including a superimposed internal-wave-induced sound-speed perturbation field, but only for rays with identifiers +137 (left panel) and +151 (right panel). The points plotted are a subset of those plotted in the lower panel of Fig. 1.

steep ray micromultipath travel time spreads is shown to be consistent with the O(10-ms) spreads seen in our simulations.

A. Micromultipaths and pulse synthesis

It is useful to assign a ray identifier $\pm M$ to each ray. Here, $\pm$ is the sign of the launch angle and M is the number of ray turning points (where p changes sign) following a ray. In Fig. 8 rays with two fixed values of ray identifier (+137 and +151) are plotted in the time–depth plane. In both cases the points plotted are a subset of those plotted in Fig. 1. It is seen that the relatively steep +137 rays have a small time spread and form one of the resolved branches of the timefront seen in Fig. 1 (similar behavior was noted in Ref. 14), while the relatively flat +151 rays have a much larger travel time spread and fall within the diffuse finale region of the arrival pattern seen in Fig. 1.

Eigenrays at a fixed depth correspond to the intersections of a horizontal line, $z=$ constant, with the curve $z(\varphi_0)$; these intersections are the roots of the equation $z-z(\varphi_0)=0$. Although the discrete samples of $z(\varphi_0)$ plotted in Figs. 6 and 7 are too sparse to reveal what should be a smooth curve, it is evident that the number of eigenrays at almost all depths is very large; because rays are predominantly chaotic, this number grows exponentially in range. In principle, eigenrays can be found using a combination of interpolation and iteration, starting with a set of discrete samples of $z(\varphi_0)$. In practice, this procedure reliably finds only those eigenrays in regions where $z(\varphi_0)$ has relatively little structure. These rays have the highest intensity and are the least chaotic. At most depths there are many eigenrays with each value of the ray identifier $\pm M$. An incomplete set of eigenrays with identifier +137, found using the procedure just described, is shown in Fig. 9. We shall refer to a set of eigenrays with the same identifier in the presence of an internal-wave-induced sound-speed perturbation as a set of micromultipaths. (This term is widely used but is rarely unambiguously defined.) What distinguishes sets of micromultipaths with small M (e.g., 137), corresponding to relatively steep rays, from those with larger M (e.g., 151), corresponding to relatively flat rays, in the AET environment is that, except near ray turning depths and accidental crossings, the former are temporally resolved while the latter are not. [Note that in order for a set of micromultipaths to be temporally resolved it must have a travel time distribution that is separated from neighboring distributions by $(\Delta f)^{-1}$ or more.]

An important (and surprising) feature of sets of micromultipaths is that they are highly nonlocal in the sense that interspersed (i.e., having intermediate launch angles) among a group of micromultipaths are rays with M values that differ by several units. We shall refer to this behavior, which is illustrated in Fig. 10, as spectral nonlocality. (Here, the term spectral is used in the sense that an integral representation of the wave field over ray-like contributions with variable launch angle is a spectral representation.) In view of the observation that, in the presence of internal waves in the AET environment, ray motion is strongly chaotic and the function $M(\varphi_0)$ has local oscillations of several units, it is remarkable that the early portion of the timefront (see Fig. 1) is not destroyed. As ray launch angle φ_0 is continuously increased, a continuous curve is traced out in (T,z) at the range of the AET receiving array. Surprisingly, for larger launch angles the curve that is traced out in (T,z) very nearly coincides with the timefront that would have been traced out in the absence of internal waves—except that in the presence of internal waves the curve includes numerous cusps at which the direction of motion along the timefront reverses. Spectral nonlocality of micromultipaths is revealed in the (T,z) curve by the many direction reversals and the associated temporary local excursions over several branches

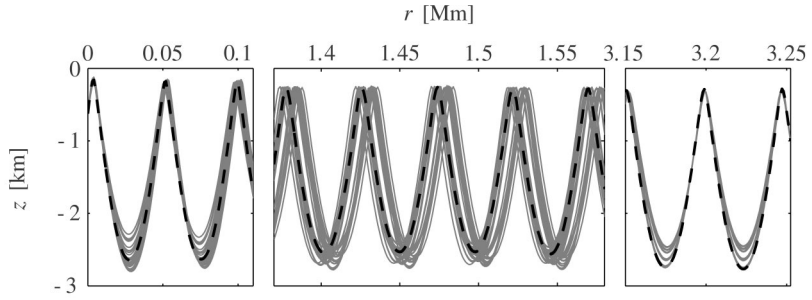


FIG. 9. Three segments of the range–depth plane along the AET transmission path showing a subset of a set of micromultipaths. The dashed line shows the path of the three rays with the largest amplitudes; these three rays are not resolved on this plot.

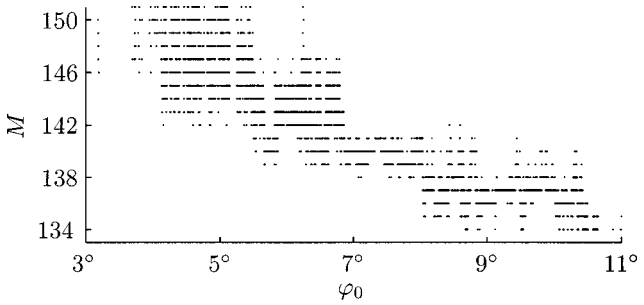


FIG. 10. Ray identifier vs launch angle for a subset of the rays that are plotted in Fig. 6.

of the timefront, each branch corresponding to a different value of *M*.

The nonlocality of a set of micromultipaths in the range–depth plane, which we shall refer to as spatial nonlocality, is seen in Fig. 9. Note that there are significant differences in the upper and lower turning depths of the plotted micromultipaths, and that these rays are spread in range by a significant fraction of a ray cycle distance. In simulations using $E=1.0\,E_{\mathrm{GM}}$, not shown, the spatial nonlocality, especially in range, of a set of micromultipaths is much stronger than that shown in Fig. 9.

The result of performing a ray-based synthesis of what appears in the measurements to be an isolated arrival is shown in Fig. 11. This involves finding a complete set of micromultipaths, including their travel times and Maslov indices, and coherently summing their contributions. Unfortunately, owing to the predominantly chaotic motion of ray trajectories in the AET environment in the presence of internal waves, it is extremely difficult to find a complete set of micromultipaths. Indeed, the constituent ray arrivals used in the Fig. 11 synthesis do not constitute a complete set of micromultipaths. This is less important than might be expected because standard eigenray finding techniques easily locate the strongest micromultipaths; the missing micromultipaths in the Fig. 11 synthesis are highly chaotic rays that have very small amplitudes. An interesting result of this synthesis is that the micromultipath-induced time spread in the synthesized pulse is only about 1 ms, which is in very good agreement with the AET measurements.[24,25] This approximately 1-ms spread is the difference in the width (approximately 25 ms) of the transmitted pulse, two cycles of a 75-Hz carrier, and the width of the coherent sum of micromultipath contributions (see Fig. 11). Note that the travel-time spread among the micromultipaths found—about 11 ms—is much greater than the resulting spread of the synthesized pulse. The principal cause of this difference in our simulations was that the travel-time spreads of the dominant micromultipaths were much smaller than the total micromultipath travel-time spreads.

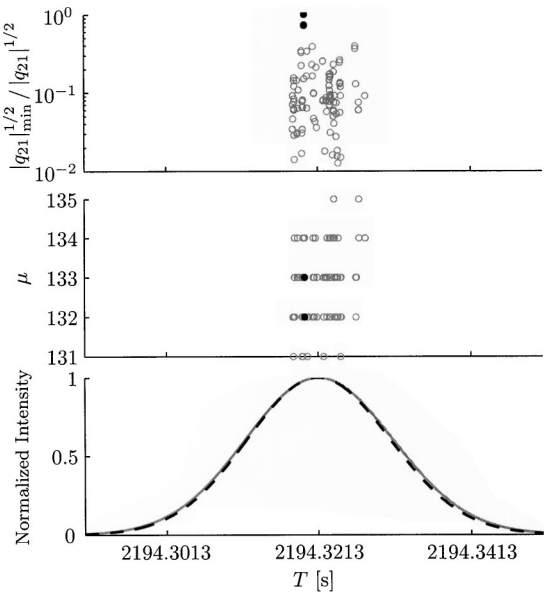


FIG. 11. Upper panel: normalized ray intensity vs travel time for a subset of the eigenrays (micromultipaths) with identifier +137 that connect the AET source and a receiver at depth 1005 m at the AET range. Middle panel: the corresponding Maslov indices vs travel time. Lower panel: envelope of the waveform synthesized by coherently adding the contributions from all the micromultipaths shown (solid curve), and the normalized envelope of the AET source waveform (dashed line). The solid circles in the upper and middle panels identify the three rays (two of the three solid circles fall on top of one another) whose intensity is greatest; open circles identify the remaining rays.

The qualitative features of micromultipath travel time, amplitude, and Maslov index distributions shown in Fig. 11 are typical of distributions found at other locations along the early portion of the timefront in our simulations using $E=0.5\,E_{\mathrm{GM}}$. Micromultipath travel-time spreads were typically 10 to 15 ms, with only a few of these micromultipaths, whose travel-time spread was typically less than 2 ms, having dominant amplitudes. One surprising aspect of this behavior is the very small travel-time spread of the dominant micromultipaths. We speculate that the explanation for this behavior is that the dominant micromultipaths are frequently organized locally by a simple caustic structure. This would account, in addition, for the observation that the dominant micromultipaths have Maslov indices that differ by no more than one unit.

Simulations (not shown) using $E=1.0\,E_{\mathrm{GM}}$ result in synthesized steep ray pulses that are spread in time much more than is shown in Fig. 11. With the stronger internal wave field, the total micromultipath travel-time spread increases by approximately a factor of 2—to about 25 ms. More importantly, however, the dominant micromultipaths have travel-time spreads comparable to the total micromultipath travel-time spread, leading to synthesized pulses that are spread by 5 to 10 ms, rather than 1 to 2 ms. Thus, measured early arrival time spreads are in better agreement with $E=0.5\,E_{\mathrm{GM}}$ simulations than with $E=1.0\,E_{\mathrm{GM}}$ simulations. The question of which value of the internal wave strength *E* in our simulations gives the best fit to the observations will be revisited in the next section.

An interesting and unexpected feature of our simulations is that the Maslov index μ was consistently lower than the number of turns *M* made by the same ray. The distribution of $M-\mu$ typically had a peak at 3 to 4 units and a width, within which all but a few outliers were contained, of 3 units in the $E=0.5\,E_{\mathrm{GM}}$ simulations (see Fig. 11); these numbers increased slightly in the $E=1.0E_{\mathrm{GM}}$ simulations. The Maslov index was computed by counting zeros of q_{21} (9) following ray paths. The quantity q_{21} was computed by solving the

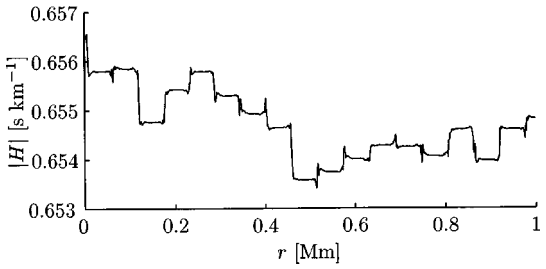


FIG. 12. $|H|$ vs range following a ray emitted on the sound channel axis with a launch angle of 11° in a canonical environment with an internal-wave-induced sound-speed perturbation field superimposed.

coupled ray-variational equations (5), (6), (7), (11), and (12). The discrepancy between M and μ in our simulations is mildly surprising because it is known[36] that in range-independent environments these numbers differ by no more than 1 unit.

B. Micromultipath travel time spreads

In the preceding subsection we saw that associated with the chaotic motion of ray trajectories is extensive micromultipathing, and that the micromultipathing process is both spectrally and spatially nonlocal. Surprisingly, on the early branches of the timefront the micromultipathing process causes only very small time spreads and does not lead to a mixing of ray identifiers. The foregoing discussion centering on Fig. 11 described how micromultipath travel-time spreads of 10 to 15 ms can be reconciled with measured wave-field time spreads of 1 to 2 ms. To complete this picture it is necessary to explain why the micromultipath travel-time spreads of the early (steep) ray arrivals are only 10 to 15 ms. We shall now provide such an explanation.

Computing time spreads is conceptually straightforward using ray methods. In a known environment one finds a large ensemble of rays with the same identifier that solves the same two-point boundary value problem; the spread in travel times over many ensembles of such rays, each ensemble computed using in a different realization of the environment, is the desired quantity. The constraint that all of the computed rays have the same fixed endpoints complicates this calculation. In the following we exploit an approximate form of the eigenray constraint that simplifies the calculation; the approximate form of the eigenray constraint is first-order accurate in a sense to be described below.

Before describing the manner in which the eigenray constraint is imposed, it is useful to describe the physical setting in which the constraint will be applied. Figure 12 shows $|H|$ vs r following a moderately steep ray (axial angle approximately 11°) in a canonical environment[37] on which the previously described [based on the measured AET $N(z)$ profile] internal-wave-induced sound-speed perturbation field was superimposed. It is seen that $H(r)$ consists of a sequence of essentially constant H segments, separated by near-step-like jumps which are caused by the sound-speed perturbation. The near-step-like jumps occur at the ray's upper turning points. A simple model that captures this behavior—the so-called apex approximation[31]—assumes that the transition regions can be approximated as step functions. This is, of course, an idealization as $H(r)$ must be a smooth function for any smooth $c(z,r)$. (The validity of the apex approximation is linked to the anisotropy and inhomogeneity of oceanic internal waves; details of the background sound-speed profile are not important. For this reason we have chosen to illustrate this effect using a background canonical profile rather than the AET environment.) Figure 13 shows a schematic diagram of two rays. One ray has undergone two internal-wave-induced apex scattering events; the other ray is an unscattered ray with the same launch angle as the scattered ray.

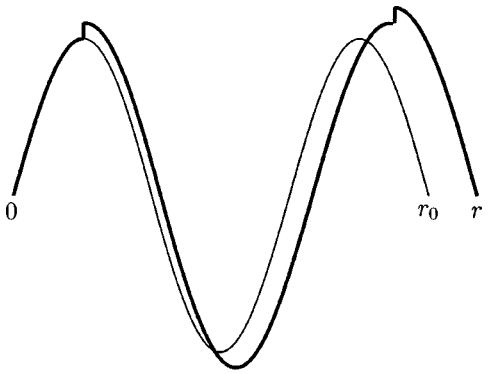


FIG. 13. Schematic diagram showing a twice apex-scattered ray (heavy line) and the path that the same ray would have followed in the absence of apex scattering events (light line).

Our strategy to build the eigenray constraint into an estimate of travel-time spreads can now be stated. We consider a scattered and an unperturbed (that is, in the absence of internal-wave-induced scattering events) ray with the same ray identifier; the rays shown schematically in Fig. 13 have identifier $+3$. In addition to having the same ray identifier, the scattered and unperturbed rays are assumed to start from the same point and end at the same depth, but they will generally end at different ranges. We assume that the scattered ray has travel time T_r and is one of many scattered eigenrays with the same ray identifier at range r. We estimate the travel time $T(r)$ of an unperturbed eigenray at range r whose ray identifier is the same as that of the scattered eigenray. It will be shown that $\delta T(r)=T_r-T(r)$ vanishes to first-order if the apex approximation is exploited. Because the same result applies to all of the scattered eigenrays, it follows that the time spread vanishes to first-order, independent of r, if the apex approximation is exploited. Note that this result is not expected to apply, even approximately, to the late near-axial rays where the apex approximation is known to fail.

The travel time $T(r)$ of the eigenray in the unperturbed ocean can, of course, be computed numerically, but this provides little help in deriving an analytical estimate of δT. Instead, consider a ray in the unperturbed ocean with the same launch angle as one of the scattered eigenrays, as shown in Fig. 13. In general, the range of this ray at the receiver depth, after making the appropriate number of turns, will be $r_0 \neq r$. But, $T(r)$ can be estimated from $T(r_0)$. This follows from the observation that ray travel time (assuming a point source initial condition, for instance) is a continuously differentiable function of z and r, with $\nabla T=\mathbf{p}$, where $\mathbf{p}$ is the ray slowness vector. Because of this property $T(z,r)$ can be

expanded in a Taylor series. If we consider rays with the same ray identifier and fix the final ray depth to coincide with the receiver depth, then

$$T(r) \approx T(r_0) + p_r(r_0)(r - r_0), \tag{14}$$

where $p_r = -H$ is the r component of the ray slowness vector. [More generally, $T(z,r)$ consists of multiple smooth sheets that are joined at cusped ridges; in spite of this complication, the property of continuous differentiability is maintained, even in a highly structured ocean. For our purposes, we need only consider one sheet of this multivalued function. Note also that our application of (14) in the background ocean makes this expression particularly simple to use.] It follows from (14) that

$$\delta T(r) = T_r - T(r) \approx T_r - T(r_0) - p_r(r_0)(r - r_0) \tag{15}$$

is, correct to first-order, the travel-time difference at range r between eigenrays with and without internal waves.

To evaluate $\delta T(r)$, we shall assume that the background environment is range independent. Also, consistent with the use of the apex approximation, we may treat the perturbed environment as piecewise range independent. Within each range-independent segment it is advantageous to make use of the action-angle variable formalism (see, e.g., Refs. 11, 22, or 38). In terms of the action-angle variables (I, θ), $\tilde{H} = \tilde{H}(I)$ and the ray equations are

$$\frac{d\theta}{dr} = \frac{\partial \tilde{H}}{\partial I} = \omega(I), \tag{16}$$

$$\frac{dI}{dr} = -\frac{\partial \tilde{H}}{\partial \theta} = 0, \tag{17}$$

and

$$\frac{dT}{dr} = I\frac{d\theta}{dr} - \tilde{H}(I) = I\omega(I) - \tilde{H}(I), \tag{18}$$

where the physical interpretation of $\tilde{H}$ as $-p_r$ is maintained. The angle variable can be defined to be zero at the upper turning depth of a ray, and increases by 2π over a ray cycle (double loop). It follows that the frequency $\omega(I) = 2\pi/R(I)$, where $R(I)$ is a ray cycle distance. Integrating the ray equations over a complete ray cycle then gives $R(I) = 2\pi/\omega(I)$ and $T(I) = 2\pi(I - \tilde{H}(I)/\omega(I))$, with I constant following each ray. In the apex approximation I jumps discontinuously at each ray's upper turning depth; for the perturbed ray $R(I + \Delta I) \approx 2\pi/\omega(I) - 2\pi\Delta I\omega'(I)/(\omega(I))^2$ and $T(I + \Delta I) \approx 2\pi(I - \tilde{H}(I)/\omega(I)) + 2\pi\Delta I\tilde{H}(I)\omega'(I)/(\omega(I))^2$, where $\omega'(I) = d\omega(I)/dI$. Note that, like these expressions, Eqs. (14) and (15) are first-order accurate in ΔI.

Consider again Eq. (15) and Fig. 13 with I equal to I_0, I_1, and I_2 in the left, center, and right ray segments, respectively. The left segment gives no contribution to δT as the ray has not yet been perturbed. In the center ray segment

$$T_r - T(r_0) \approx 2\pi(I_1 - I_0)\frac{\tilde{H}(I_0)\omega'(I_0)}{(\omega(I_0))^2}, \tag{19}$$

and

$$-p_r(r_0)(r - r_0) \approx \tilde{H}(I_0)\left[-2\pi(I_1 - I_0)\frac{\omega'(I_0)}{(\omega(I_0))^2}\right]. \tag{20}$$

These terms are seen to cancel. Note that if additional complete cycle ray segments are added to the center section of the ray, Eqs. (19) and (20) are unchanged except that $I_1 - I_0 = \Delta I_1$ is replaced by $\Sigma_{i=1}^{n}\Delta I_i = I_n - I_0$; again, the two terms cancel. In the final (incomplete cycle) ray segment the difference between the terms $T_r - T(r_0)$ and $p_r(r_0)(r - r_0)$ can be shown to be $O((\Delta I)^2)$; the terms do not exactly cancel because the θ values of the perturbed and unperturbed rays are generally not identical at the receiver depth. Thus, to first-order in ΔI, $\delta T = 0$, independent of range, if the apex approximation is valid. This simple calculation provides an explanation of why, in spite of extensive ray chaos, the time spreads of the early AET ray arrivals are quite small.

Several comments concerning the preceding calculation are noteworthy. First, we note that although it was assumed that the background sound-speed structure is range independent, the preceding argument also holds in the presence of slow background range dependence, i.e., with structure whose horizontal scales are large relative to a typical ray double-loop length. Adiabatic invariance in such environments guarantees that while H is not constant following rays between apex scattering events, I is nearly constant. Second, after n random kicks, $I_1 - I_0$ in (19) and (20) is replaced by $I_n - I_0 \approx \sqrt{n}(\Delta I)_{\mathrm{rms}}$, and the magnitude of (19) under AET conditions is $O(1\ \mathrm{s})$. This is an example of a travel time-spread estimate that fails to enforce the eigenray constraint. This calculation shows that the difference between travel-time estimates that do and do not enforce the eigenray constraint can be quite significant. (The constrained estimate is refined below.) Third, it should be emphasized that Eqs. (19) and (20) apply (approximately) only to the steep rays because the apex approximation applies only to the steep rays. We have not addressed the time spreads of near-axial rays. Note, however, that Fig. 8 shows that time spreads are larger for flatter rays. Fourth, the above calculation shows that, to lowest order in ΔI, there is no internal-wave-scattering-induced travel-time bias if the apex approximation is strictly applied. And fifth, the arguments leading to Eqs. (19) and (20) apply whether the scattered rays are chaotic or not.

A relaxed form of the apex approximation in which the action jump transition region has width $\Delta\theta$ gives a nonzero travel-time-spread estimate. In the transition region, taken for convenience to be $0 \leqslant \theta \leqslant \Delta\theta$, one may choose a Hamiltonian of the form $\tilde{H} = \tilde{H}(I - s\theta)$, where $s = \Delta I/\Delta\theta$, and $\tilde{H} = \tilde{H}(I)$ elsewhere. It follows that in the transition region $I(\theta) = I_0 + s\theta$ and $I =$ constant elsewhere. A simple generalization of the above calculation then gives for a complete ray cycle

$$T_r - T(r_0) \approx (2\pi - \Delta\theta)(I_1 - I_0)\frac{\tilde{H}(I_0)\omega'(I_0)}{(\omega(I_0))^2} + \frac{\Delta\theta}{2}(I_1 - I_0), \tag{21}$$

and

$$-p_r(r_0)(r-r_0)\approx\tilde{H}(I_0)\left[-(2\pi-\Delta\theta)(I_1-I_0)\frac{\omega'(I_0)}{(\omega(I_0))^2}\right], \tag{22}$$

so the sum [recall Eq. (15)] is

$$\delta T\approx\frac{\Delta\theta}{2}(I_1-I_0). \tag{23}$$

It should be emphasized that this expression is first-order accurate in $\Delta I=I_1-I_0$, but that no assumption about the smallness of $\Delta\theta$ has been made. As noted above, incomplete cycle end segment pieces give $O((\Delta I)^2)$ corrections to δT. If $\Delta\theta$ has approximately the same value at each upper turn, then one has, after n upper turns, correct to $O(\Delta I)$

$$\delta T\approx\frac{\Delta\theta}{2}\sum_{i=1}^{n}(I_i-I_{i-1})=\frac{\Delta\theta}{2}(I_n-I_0)$$

$$\approx\frac{\Delta\theta}{2}\sqrt{n}(\Delta I)_{\mathrm{rms}}. \tag{24}$$

Consistent with the numerical simulations shown in Fig. 12, $\Delta\theta\approx0.8$ radians and $\Delta I\approx4$ ms. (We, and independently F. Henyey [personal communication, 2002], have confirmed that these estimates also apply under AET-like conditons.) With these numbers and $\sqrt{n}=8$, appropriate for AET, (24) gives a time-spread estimate of approximately 13 ms, in approximate agreement with the numerical simulations shown in Fig. 11. Note that (24) does not preclude a travel-time bias. A cautionary remark concerning the use of (24) is that Virovlyansky[38,39] has pointed out that, owing to secular growth, the $O((\Delta I)^2)$ contribution to δT may dominate the $O(\Delta I)$ contribution at long range.

Figure 8 shows that in the AET environment simulated near-axial ray time spreads are greater than simulated steep ray time spreads. We do not fully understand this behavior. A possible explanation for this behavior is that time spreads increase as rays become increasingly flat owing to the breakdown of the apex approximation. But, one would expect that this trend should be offset, in part or whole, by the relative smallness of internal-wave-induced sound-speed perturbations near the sound-channel axis. In addition, we have seen some evidence that an additional factor may be important in the AET environment. Namely, we have observed a positive correlation between travel-time spreads and stability exponents; stability exponents in the AET environment are shown in Fig. 5. We have chosen not to dwell on flat ray time spreads in this paper because there are no AET measurements of these spreads to which simulations can be compared. It is clear, however, that the issues just raised need to be better understood.

IV. WAVE-FIELD INTENSITY STATISTICS

In this section we consider the statistical distribution of the intensities of both the early and late AET arrivals. Recall that experimentally the early arrival intensities have been shown[24,25] to approximately fit a lognormal probability density function (PDF), and the late arrival intensities have been shown[25] to fit an exponential distribution. The late arrival exponential distribution is not surprising as this distribution is characteristic of saturated statistics. The early arrival near-lognormal distribution is surprising, however, inasmuch as theory[31,24] predicts saturated statistics, i.e., an exponential intensity PDF. It has been argued[40] that the theory can be modified in such a way as to move the early arrival prediction from saturated to unsaturated statistics. The latter regime is characterized by a lognormal intensity PDF. This fix is conceptually problematic inasmuch as, in this theory, the unsaturated regime is characterized by the absence of micromultipaths, which seriously conflicts with the numerical simulations presented in the previous section where the number of micromultipaths is very large. In this section we provide self-consistent explanations for both the late arrival exponential distribution and the early arrival lognormal distribution. The challenge is to reconcile the early arrival near-lognormal intensity PDF with the presence of a large number of micromultipaths. Some of the arguments presented are heuristic, and some build on the results of numerical simulations. A complete theoretical understanding of intensity statistics has proven difficult.

Our approach to describing wave-field intensity statistics builds on the semiclassical construction described by Eq. (2). At a fixed location it is seen that the wave-field amplitude distribution is determined by the distribution of ray amplitudes and their relative phases. Note that both travel times and Maslov indices influence the phases of ray arrivals. Complexities associated with transient wave fields and caustic corrections will be discussed below.

An important observation is that in the AET environment, including internal-wave-induced sound-speed perturbations, simulated geometric amplitudes of both steep and flat rays approximately fit lognormal PDFs. This is shown in Fig. 14. Previously, it has been shown[20] that ray intensities in a very different chaotic system also fit a lognormal PDF; in that system single-scale isotropic fluctuations are superimposed on a homogeneous background. (Note that all powers of a lognormally distributed variable are also lognormally distributed, so ray amplitudes have this property if and only if ray intensities have this property.) The apparent generality of the near-lognormal ray intensity PDF suggests that it applies generally to ray systems that are far from integrable; the arguments presented in Ref. 20 suggest that this should be the case.

In the two subsections that follow, the intensity distributions of the early and late AET arrivals, corresponding to steep and flat rays, respectively, will be discussed separately. The phase assumptions made in these subsections limit the anticipated validity of the results presented to ''interior'' portions of the early resolved timefront branches and the finale region, respectively. Thus, accidental crossings of different resolved branches of the timefront are excluded from consideration, as are regions near turning points where adjacent timefront branches merge. Similarly, the narrow transition region separating the early resolved branches of the timefront from the finale region is excluded from consideration. These restrictions are consistent with the manner in which the AET intensity statistics were computed.[24,25] Also, it should be

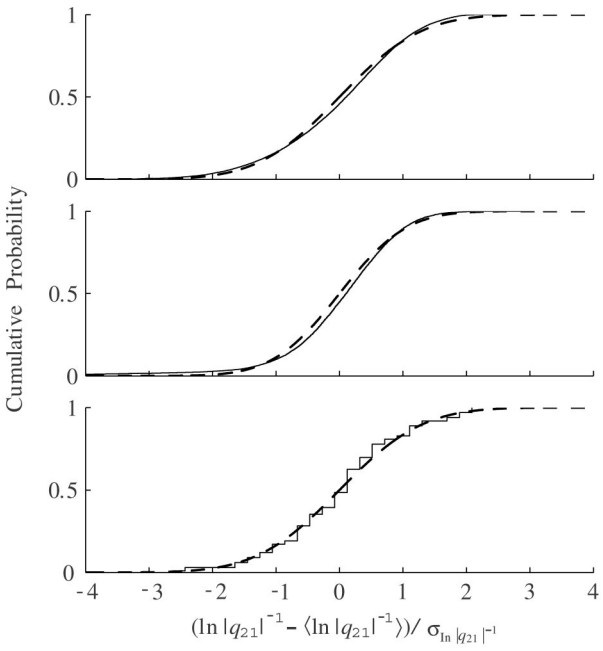


FIG. 14. Cumulative probability density of ray intensity in the AET environment including a superimposed internal-wave-induced sound-speed perturbation field (solid lines). Upper panel: flat rays ($|\varphi| \leqslant 5°$) sampled uniformly in launch angle. Middle panel: steep rays ($6° \leqslant |\varphi| \leqslant 11°$) sampled uniformly in launch angle. Lower panel: eigenrays with +137 identifier connecting the AET source and a receiver at depth 1005 m at the AET range. In all three panels the dashed lines correspond to lognormal distributions whose mean and variance approximately match those of the simulations.

noted that our implicit assumption that intensity statistics should be independent of position (T,z) in each of these interior regions is consistent with both our simulations and the observations.[24,25]

A. Early arrivals

We saw in the previous section that the micromultipaths that make up one of the early arrivals have the same ray identifier and have a very small spread in travel time. The dominant micromultipaths were seen (see Fig. 11) to have a time spread of approximately 1 ms, which is a small fraction of the approximately 13-ms period of the 75-Hz carrier wave. Also, our simulations show that the Maslov indices of the dominant micromultipaths differ by no more than 1 unit. These conditions dictate that interference among the dominant micromultipaths is predominantly constructive. Because travel-time differences are so small, the pulse shape should have negligible influence on the distribution of peak intensities. In a model of the early AET arrivals consisting of a superposition of interfering micromultipaths, the micromultipath properties that play a critical role in controlling peak wave-field intensities are thus: (1) their amplitudes have a near-lognormal distribution; (2) the dominant micromultipaths have negligible travel-time differences; and (3) the dominant micromultipaths have Maslov indices that differ by no more than 1 unit. It should be noted that very different behavior would have been observed if: the source bandwidth were significantly more narrow, as this would have caused micromultipaths with different ray identifiers to interfere with one another; or the source center frequency were significantly higher as phase differences between interfering micromultipaths would then have been significant.

An additional subtlety must be introduced now: the PDF of the intensities of the constituent micromultipaths that make up a single arrival is not identical to the PDF described in Ref. 20 and shown in the upper and middle panels of Fig. 14. The latter PDF describes the distribution of intensities of randomly (with uniform probability) selected rays leaving the source within some small angular band and whose range is fixed. This PDF is biased in the sense that it overcounts the micromultipaths with large intensities and undercounts those with small intensities. Unbiased micromultipath intensity PDFs can be constructed from the biased PDFs shown. To do so, consider a manifold (a smooth curve in phase space corresponding to a fan of initial rays) which begins near (z_0,p_0) and arrives in the neighborhood of (z_r,p_r) at range r. Sampling in fixed steps of the differential δp_0 (as was done to produce the upper and middle panels of Fig. 14; this is equivalent to uniform random sampling) leads to a highly nonuniform density of points on the final manifold, since $(z_0,p_0+\delta p_0)$ propagates to $(z_r+q_{21}\delta p_0, p_r+q_{11}\delta p_0)$. The greater q_{21}, the lower the density of points locally at final range. A uniform sampling in final position is achieved instead by considering the initial conditions that would lead to $(z_r+\delta z_r, p_r+q_{11}/q_{21}\delta z_r)$. Its density of points on the initial manifold can be deduced from its initial condition, $(z_0, p_0 + \delta z_r/q_{21})$. It is necessary to sample q_{21} times more densely on the initial manifold in order to achieve uniform sampling in δz_r at range r. To account for this effect, we need to know the PDF for q_{21} with uniform initial sampling. Roughly speaking, the PDF of the absolute values of the individual matrix elements of Q have the same form as for $|\mathrm{Tr}(Q)|$, apart from a shift of the centroid that is lower-order in range than the leading term. From the results of Ref. 20, the (biased in the sense described above) probability that q_{21} falls in the interval between x and $x+dx$ is

$$\rho_{|q_{21}|}(x) = \sqrt{\frac{1}{2\pi r(\bar{\nu}-\nu_L)}}\frac{1}{x}\exp\left[\frac{-(\ln(x)-\nu_L r)^2}{2r(\bar{\nu}-\nu_L)}\right], \quad x \geqslant 0. \tag{25}$$

Here, ν_L is the true Lyapunov exponent, and $\bar{\nu}$ is equal to ν_L plus an additional contribution due to the fluctuations in the values of q_{21} over an ensemble of rays. The new (unbiased in the sense described above) PDF, ρ', for uniform sampling in δz_r is related to the previous one by

$$\rho'_{|q_{21}|}(x) = \frac{x}{\langle x\rangle}\rho_{|q_{21}|}(x) = \sqrt{\frac{1}{2\pi r(\bar{\nu}-\nu_L)}}\frac{1}{x}\exp\left[\frac{-(\ln(x)-\bar{\nu}r)^2}{2r(\bar{\nu}-\nu_L)}\right], \quad x \geqslant 0, \tag{26}$$

where the factor x accounts for the extra counting weight of q_{21}, and $\langle x\rangle$ just preserves normalization and is calculated using Eq. (25). This calculation shows that the unbiased micromultipath ray intensity PDF also has a lognormal distribution; the only change relative to the biased PDF is an

increase in the mean from $\nu_L r$ to $\overline{\nu r}$. Because lognormality is maintained, this correction represents only a trivial change to the problem.

Approximate lognormality of the constrained (eigenray) PDF of ray intensity is shown in the lower panel of Fig. 14. This PDF was constructed using the same eigenrays that were used to produce Figs. 9 and 11. The corresponding unconstrained ray intensity PDF is shown in the middle panel of Fig. 14. (Two constraints—fixed receiver depth and fixed ray identifier—are built into the lower panel PDF. A very similar constrained PDF results if only the receiver depth constraint is applied, provided ray launch angles are limited to the ''steep'' ray band used to construct the middle panel.) It should be noted that the constrained (eigenray) PDF shown in the lower panel of Fig. 14 was constructed from numerically found eigenrays; because weak eigenrays are difficult to find numerically they are undercounted and the PDF is biased. Because this bias is confined to the tail of the distribution corresponding to the very weak eigenrays, whose contribution to the wave field is expected to be small, it is not expected to alter the argument that follows—which assumes only that micromultipath amplitudes are near-lognormally distributed.

We return now to the problem of simulating the early AET arrivals. It is tempting to think that because the constituent micromultipath amplitudes have a near-lognormal distribution, the sum of the micromultipath contributions should also be near-lognormally distributed. Unfortunately, this is generally not the case. Consider, for example, the special case in which phase, including Maslov index, differences are negligible. Then, all micromultipaths interfere constructively and peak wave-field amplitudes can be modeled as the sum of many lognormally distributed variables. Because all moments of the lognormal distribution are finite, the central limit theorem applies. Under these conditions, if sufficiently many contributions are summed, the distribution of the sums—the wave-field amplitudes—would be a Gaussian.

To simulate the statistics of the early AET arrivals (recall Fig. 14 and the accompanying discussion), we have used several variations of a simple model. An arrival was modeled as a sum of n_m interfering micromultipaths whose: (1) amplitudes are lognormally distributed; and (2) phases, $\sigma T_i - \mu_i \pi/2 \bmod 2\pi$, have a PDF with a clearly identifiable peak. Micromultipath contributions were coherently added. The peak intensity of the sum—whose travel time is not known *a priori*—was then recorded. Using an ensemble of 10^4 peak intensities, a peak-intensity PDF was then constructed. Peak-intensity PDFs constructed in this fashion were found to be very close to lognormal; a typical example is shown in Fig. 15. In this example $f=75$ Hz, the T_i's were identical, $\mu_i \in \{j, j+1\}$ with equal probability (note that choice of the integer j is unimportant), and $n_m=5$. Other combinations of distributions for T_i (either a Gaussian or the limiting case of a delta distribution), μ_i (taken either from $\{j, j+1\}$ or $\{j-1, j, j+1\}$ with equal probability), and the parameter n_m (between 2 and 100) were tested. These simulations showed that, provided the phase constraint noted above was satisfied, a near-lognormal peak-intensity PDF resulted.

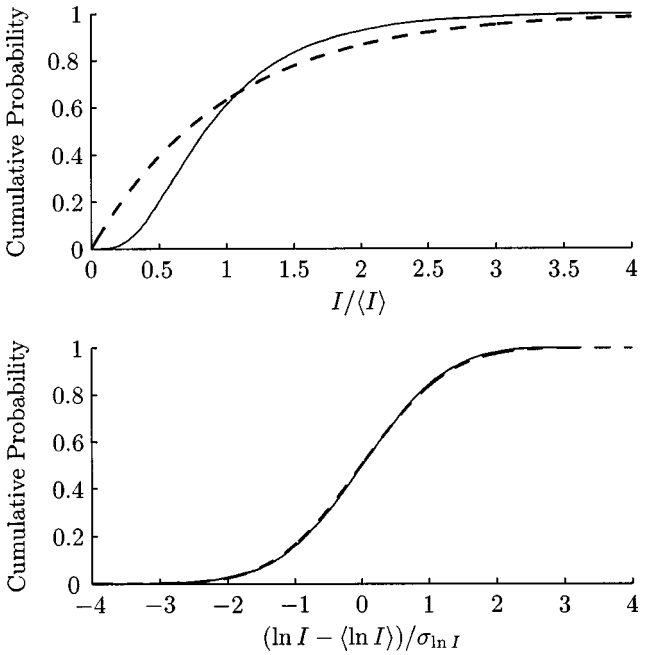


FIG. 15. Cumulative probability density of peak wave-field intensity simulated using a simple model as described in the text (solid lines). The same simulated cumulative density is compared to exponential (upper panel dashed line) and lognormal (lower panel dashed line) cumulative density functions. This figure should be compared to Fig. 14 in Ref. 24.

Two points regarding this simple model are noteworthy. First, this model does not constitute a theory of wave-field peak-intensity statistics, but it does serve to demonstrate that our ray-based simulations of the early AET arrivals are consistent with the measured distribution of peak intensities. Second, simulations (not shown), performed with $E = 1.0\, E_{\mathrm{GM}}$ yield sets of dominant micromultipaths that violate assumption (2); phases are uniformly distributed and summing micromultipath contributions yields a distribution of peak intensities that is not close to lognormal. Thus, our simulations suggest that a near-lognormal distribution of early arrival peak intensity requires a relatively weak internal wave field.

A complication not accounted for in the preceding discussion is the presence of caustics. At caustics geometric amplitudes (8) diverge and diffractive corrections must be applied. At short range (on the order of the first focal distance—a few tens of km in deep ocean environments) we expect that intensity fluctuations will be dominated by diffractive effects. The entire wave field should be organized by certain high-order caustics which leads to a PDF of wave-field intensity with long tails.[41–43] In spite of the importance of diffractive effects at short range—and probably also at very long range—we believe that, in the transitional regime described above, intensity fluctuations are not dominated by diffractive effects. This somewhat counterintuitive behavior can be understood by noting that in the vicinity of caustics the importance of diffractive corrections to (8) decreases as the curvature of the caustic increases. A quantitative measure of caustic curvature is the second derivative of the Lagrangian manifold (recall Fig. 4) $z''(p)$ at a local extremum.

(Higher-order caustics can be treated using an extension of the same argument.) Under chaotic conditions the curvature of caustics increases, on average, with increasing range, as the Lagrangian manifold develops structure on smaller and smaller scales. Thus, the fraction of the total number of multipaths that require caustic corrections decreases, on average, with increasing range. This is true even as the number of caustics grows exponentially, on average, in range. This argument leads to the somewhat paradoxical conclusion that, prior to saturation at least, we expect that the importance of caustic corrections decreases with increasing range.

B. Late arrivals

Recall that in the finale region of the AET wave fields, sets of micromultipaths with different ray identifiers, all corresponding to near-axial rays, are not temporally resolved. At each (T,z) in the finale region the wave field can be modeled as a superposition of micromultipath contributions with random phases. The quadrature components of the wave field at each (T,z) have the form of sums of terms of the form

$$x_i = a_i \cos(\phi_i), \tag{27}$$

and

$$y_i = a_i \sin(\phi_i), \tag{28}$$

where ϕ_i is a random variable uniformly distributed on $[0,2\pi)$. The distribution of a_i is close to lognormal, but a correction must be applied to account for pulse shape, as many of the interfering micromultipaths partially overlap in time. This correction is unimportant inasmuch as the central limit theorem guarantees that, provided the distributions of x_i and y_i have finite moments, the distributions of sums of x_i and y_i converge to zero mean Gaussians. Thus, wave-field intensity is expected to have an exponential distribution, consistent with the observations. The comments made earlier about caustics apply here as well.

C. Intensity statistics discussion

The question of what causes the transition from the structured early portion of the AET arrival pattern to the unstructured finale region deserves further discussion. Recall that the early resolved arrivals have small time spreads and peak intensities that are near-lognormally distributed, while the finale region is characterized by unresolved arrivals and near-exponentially distributed intensities. In both regions the time spreads and intensity statistics are consistent with each other inasmuch as in our simulations a near-lognormal intensity distribution is obtained only when there is a preferred phase, while the exponential distribution is generated when phases are random, i.e., when phases are uniformly distributed on $[0,2\pi)$.

With these comments in mind, it is evident that the most important factor in causing the transition to the finale region is the increase in internal-wave-scattering-induced time spreads as rays become less steep; as time spreads increase, neighboring timefront branches blend together and the phases of interfering micromultipaths get randomized. The trend toward increasing time spreads as rays become less steep is evident in Fig. 8. The surprising result is that the scattering-induced time spreads of the steep arrivals are so small; we have seen that this can be explained by making use of the apex approximation. As noted at the end of the previous section, we do not fully understand the cause of the trend toward larger time spreads as rays flatten. In the finale region internal-wave-scattering-induced time spreads exceed the time difference between neighboring timefront branches that would have been observed in the absence of internal waves. Figure 1 shows that these time gaps decrease as rays become increasingly flat. Indeed, this figure shows that even in the absence of internal waves there would not have been any resolvable timefront branches in the last half-second or so of the arrival pattern. Some loss of temporal resolution in the measurements is of course due to the finite source bandwidth, but without internal waves (or some other type of ocean fluctuations) wave-field phases would not be random; instead, a stable interference pattern would be observed.

Finally, we note that an interesting feature of our ray simulations is that the near-axial rays have much higher stability exponents [typically about $(100\ \text{km})^{-1}$] than the steeper rays [typically about $(300\ \text{km})^{-1}$]; see Fig. 5. Also, note that Figs. 6 and 7 strongly suggest that the near-axial rays in the AET environment are much more chaotic than the steeper rays. The likely cause of the relative lack of stability of the near-axial rays in the AET environment is the background sound-speed structure. This topic is discussed in detail in Ref. 44. There it is shown that ray stability is largely controlled by a property of the background (which is assumed to be range independent) sound-speed profile, $\alpha = (I/\omega)d\omega/dI$; ray instability increases on average with increasing $|\alpha|$. Here, I is the ray action variable and $2\pi/\omega(I)$ is the ray cycle (double loop) distance; recall Eqs. (16)–(18). The dependence of α on I can be replaced by dependence on axial ray angle φ_{ax} because, in an environment with a single sound-speed minimum, I is a monotonically increasing function of φ_{ax}. Plots of α vs φ_{ax} in five 650-km block range-averaged sections of the AET environment are shown in Fig. 16. (The choice of averaging over 650-km blocks in range is, of course, arbitrary. Range averaging should be done locally, however, because the local sound-speed structure may be very different than that which results after averaging over the entire propagation path.) The relatively large near-axial ray values of $|\alpha|$ seen in this figure are consistent with the strongly chaotic nature of these rays seen in Figs. 5, 6, and 7. Unfortunately, the measured wave-field intensity statistics in the AET finale region are not very sensitive to the near-axial ray intensity PDF; as noted above, the argument leading to the expectation that wave-field intensity in the finale region should have an exponential distribution holds for a very large class of ray intensity distributions. Thus, we are not aware of any way that the AET measurements can be used to test our finding that the near-axial rays have larger stability exponents, on average, than the steeper rays.

In summary, we believe that the observed near-lognormal PDF of wave-field intensity for the early resolved AET arrivals is transitional between fluctuations dominated by caustics at short range and saturation at long range, where phase differences will be larger. Surprisingly, phase differ-

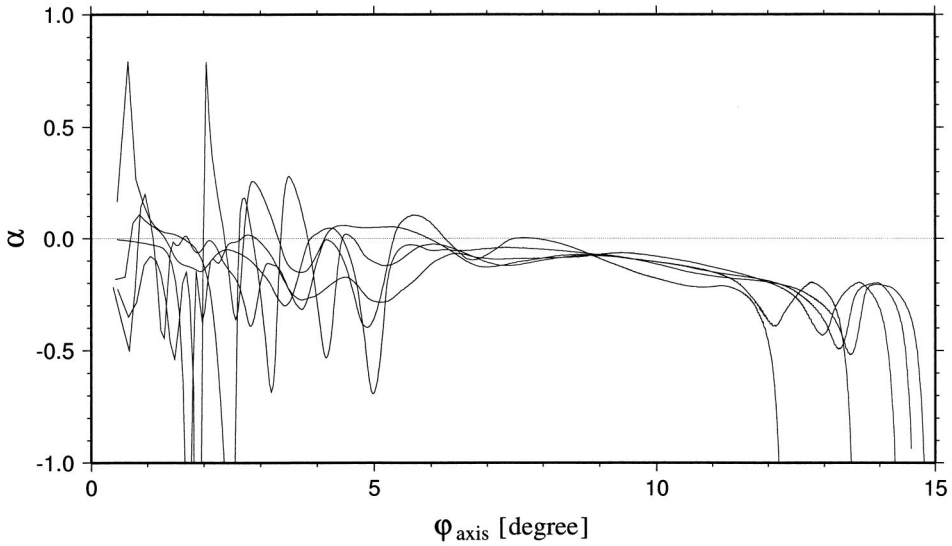


FIG. 16. Stability parameter α vs axial ray angle in five 650 km block range-averages of the AET environment.

ences among the micromultipaths that form the early resolved AET arrivals are very small so that saturation has not been reached. The late-arriving AET energy, on the other hand, is characterized by interfering micromultipaths with random phases, leading to an exponential PDF of wave-field intensity. Here, the underlying near-lognormal PDF of ray intensity is obscured. The arguments that have been presented to explain the intensity fluctuations of both the early and late AET arrivals assume that diffractive effects do not play a dominant role. Although our theoretical understanding of many of the issues raised in this section is clearly incomplete, it is worth emphasizing that our ray-based numerical simulations of intensity statistics, in which ray trajectories are predominantly chaotic, are in good qualitative agreement with the AET measurements. We do not believe that this agreement is accidental.

V. DISCUSSION AND SUMMARY

In this paper we have seen that in the AET environment, including internal-wave-induced sound-speed perturbations, ray trajectories are predominantly chaotic. In spite of extensive ray chaos, many features of ray-based wave-field simulations were shown to be both stable and in good agreement with the AET measurements. Simulated and measured spreads of acoustic energy in time, depth, and angle were shown to be in good agreement. It was shown that associated with the chaotic motion of ray trajectories is extensive micromultipathing, and that the micromultipathing process is highly nonlocal; the many micromultipaths that add at the receiver to produce what appears to be a single arrival may sample the ocean very differently. It was shown, surprisingly, that on the early arrival branches the nonlocal micromultipathing process causes only very small time spreads and does not lead to a mixing of ray identifiers. A quantitative explanation for the cause of the very small time spreads of the early ray arrivals was presented. Partially heuristic explanations for the near-lognormal and exponential distributions of wave-field intensities for the early and late arrivals, respectively, were provided.

The fact that one is able to accurately simulate many wave-field features using ray methods under conditions in which ray trajectories are predominantly chaotic may surprise some readers. Chaotic motion is, after all, intimately linked with unpredictability. This apparent paradox is reconciled by noting that while individual chaotic ray trajectories are unpredictable beyond some short predictability horizon, distributions of chaotic ray trajectories may be quite stable and have robust properties.[45] Consider, for example, the motion of a ray whose initial depth z_0 and angle φ_0 are known in the AET environment, including a known internal-wave-induced sound-speed perturbation. At the range of the AET receiving array the depth z, angle φ, travel time T, and even the number of turns M made by this ray are likely to be, for all practical purposes, unpredictable. In contrast, at the same range the distribution in (z,φ,T,M) of a large ensemble of rays with the same initial depth and whose launch angles lie within a 1° band centered on φ_0 is very stable in the sense that essentially the same distribution is seen for any large ensemble of randomly chosen (with uniform probability) rays inside this initial angular band. It is the latter property that allows us to make meaningful predictions using an ensemble of chaotic rays in a particular realization of the internal wave field. In addition, ray distributions with similar statistical properties are observed using different realizations of the internal wave field, suggesting that these statistical properties of rays are robust.

Our exploitation of results that relate to ray dynamics, including ray chaos, leads to a blurring of the distinction that is traditionally made between deterministic and stochastic wave propagation problems. The ray dynamics approach emphasizes the distinction between integrable and nonintegrable ray systems, corresponding to range-independent and range-dependent environments, respectively. In generic range-dependent environments at least some ray trajectories will exhibit chaotic motion. The oceanographic origin—mesoscale variability, internal waves, or something else—of the range-dependent structure is not critical. An important conceptual insight that can come only from exploitation of

results relating to ray chaos is that generically phase space is partitioned into chaotic and nonchaotic regions. This mixture contributes to a combination of limited determinism and constrained stochasticity.

For those who wish to exploit elements of determinism in the propagation physics for the purpose of performing deterministic tomographic inverses, the results that we have presented have important implications. We have seen, for example, that, in spite of extensive ray chaos, many ray-based wave-field descriptors are stable and predictable, and should be invertible. A less encouraging but important observation is that although the travel times of the steep early arrivals are stable and can be inverted, the spatial nonlocality of the micromultipaths limits one's ability to invert for range-dependent ocean structure, as was first pointed out in Ref. 8.

For those who wish to understand and predict wave-field statistics, our results also have important implications. These comments are based on our analysis of the AET measurements, but are expected to apply to a large class of long-range propagation problems. First, we have seen that stable and unstable ray trajectories coexist and that this influences wave-field intensity statistics. Second, we have seen that micromultipathing is extensive and highly nonlocal. The strongly nonlocal nature of the micromultipathing process is significant because: (a) this process cannot be modeled using a perturbation analysis which assumes the existence of an isolated background ray path; and (b) the effects of this process will not, in general, be eliminated as a result of local smoothing by finite frequency effects. Third, we have seen that the appearance of stochastic effects is not entirely due to internal waves. We have seen, for example, that, even in the absence of internal waves, near-axial rays in the AET environment are chaotic. Recall that we noted in the Introduction that part of the motivation for the present study was to understand the reasons underlying the finding of Colosi *et al.*[24] that the theory described in Ref. 31 fails to correctly predict the AET wave-field statistics. This failure is not surprising in view of the observation that this theory is based on assumptions that either explicitly violate, or lead to the violation of, all three of the aforementioned properties of the scattered wave field. Note also in this regard that the combination of extensive micromultipathing, small time spreads, and lognormally distributed intensities that characterizes the early AET arrivals is not consistent with any of the three propagation regimes identified in this theory.

Although we have argued that a ray-based wave-field description which relies heavily on results relating to ray chaos can account for all of the important features of the AET measurements, it should be emphasized that our analysis has some shortcomings. Recall that our simulations using $E=0.5\,E_{\mathrm{GM}}$ resulted in steep arrival time spreads and peak intensity statistics in good agreement with measurements, but that simulations using $E=1.0\,E_{\mathrm{GM}}$ did not, although there are indications from direct measurements that $E=1.0\,E_{\mathrm{GM}}$ is a more appropriate choice. We have not provided a rigorous argument to establish the connection between a near-lognormal ray intensity PDF and a near-lognormal wave-field intensity PDF. Indeed, a complete theory of wave-field statistics which accounts for complexities associated with ray chaos is lacking. This is closely linked to the subject of wave chaos,[22] which is currently not well understood. Also, we have only briefly discussed the spreads of energy in depth and angle, and more work needs to be done on quantifying time spreads.

Finally, we wish to remark that the importance of ray methods is not diminished by recent advances, both theoretical and computational, in the development of full wave models, such as those based on parabolic equations. The latter are indispensable computational tools in many applications. In contrast, the principal virtue of ray methods is that they provide insight into the underlying wave propagation physics that is difficult, if not impossible, to obtain by any other means. The results presented in this paper illustrate this statement.

ACKNOWLEDGMENTS

We thank Fred Tappert, Frank Henyey, and Dave Palmer for the benefit of discussions on many of the topics included in this paper, and the ATOC group (A. B. Baggeroer, T. G. Birdsall, C. Clark, B. D. Cornuelle, D. Costa, B. D. Dushaw, M. A. Dzieciuch, A. M. G. Forbes, B. M. Howe, D. Menemenlis, J. A. Mercer, K. Metzger, W. Munk, R. C. Spindel, P. F. Worcester, and C. Wunsch) for giving us access to the AET acoustic and environmental measurements. This work was supported by Code 321 OA of the U.S. Office of Naval Research.

[1] S. S. Abdullaev and G. M. Zaslavsky, "Stochastic instability of rays and the speckle structure of the field in inhomogeneous media," Zh. Eksp. Teor. Fiz. **87**, 763–775 (1984) [Sov. Phys. JETP **60**, 435–441 (1985)].
[2] D. R. Palmer, M. G. Brown, F. D. Tappert, and H. F. Bezdek, "Classical chaos in nonseparable wave propagation problems," Geophys. Res. Lett. **15**, 569–572 (1988).
[3] S. S. Abdullaev and G. M. Zaslavsky, "Fractals and ray dynamics in longitudinally inhomogeneous media," Sov. Phys. Acoust. **34**, 334–336 (1989).
[4] S. S. Abdullaev and G. M. Zaslavskii, "Classical nonlinear dynamics and chaos of rays in wave propagation problems in inhomogeneous media," Usp. Fiz. Nauk **161**, 1–43 (1991).
[5] M. G. Brown, F. D. Tappert, and G. Goñi, "An investigation of sound ray dynamics in the ocean volume using an area-preserving mapping," Wave Motion **14**, 93–99 (1991).
[6] F. D. Tappert, M. G. Brown, and G. Goñi, "Weak chaos in an area-preserving mapping for sound ray propagation," Phys. Lett. A **153**, 181–185 (1991).
[7] M. G. Brown, F. D. Tappert, G. Goñi, and K. B. Smith, "Chaos in underwater acoustics," in *Ocean Variability and Acoustic Propagation*, edited by J. Potter and A. Warn-Varnas (Kluwer Academic, Dordrecht, 1991), pp. 139–160.
[8] D. R. Palmer, T. M. Georges, and R. M. Jones, "Classical chaos and the sensitivity of the acoustic field to small-scale ocean structure," Comput. Phys. Commun. **65**, 219–223 (1991).
[9] K. B. Smith, M. G. Brown, and F. D. Tappert, "Ray chaos in underwater acoustics," J. Acoust. Soc. Am. **91**, 1939–1949 (1992).
[10] K. B. Smith, M. G. Brown, and F. D. Tappert, "Acoustic ray chaos induced by mesocsale ocean structure," J. Acoust. Soc. Am. **91**, 1950–1959 (1992).
[11] S. S. Abdullaev, *Chaos and Dynamics of Rays in Waveguide Media*, edited by G. Zaslavsky (Gordon and Breach, New York, 1993).
[12] F. D. Tappert and X. Tang, "Ray chaos and eigenrays," J. Acoust. Soc. Am. **99**, 185–195 (1996).
[13] G. M. Zaslavsky and S. S. Abdullaev, "Chaotic transmission of waves and 'cooling' of signals," Chaos **7**, 182–186 (1997).
[14] J. Simmen, S. M. Flatté, and G.-Yu. Wang, "Wavefront folding, chaos and

diffraction for sound propagation through ocean internal waves,'' J. Acoust. Soc. Am. **102**, 239–255 (1997).
[15] M. G. Brown, ''Phase space structure and fractal trajectories in 1 1/2 degree of freedom Hamiltonian systems whose time dependence is quasi-periodic,'' Nonlinear Process. Geophys. **5**, 69–74 (1998).
[16] M. Wiercigroch, M. Badiey, J. Simmen, and A. H.-D. Cheng, ''Nonlinear dynamics of underwater acoustics,'' J. Sound Vib. **220**, 771–786 (1999).
[17] B. Sundaram and G. M. Zaslavsky, ''Wave analysis of ray chaos in underwater acoustics,'' Chaos **9**, 483–492 (1999).
[18] M. A. Wolfson and F. D. Tappert, ''Study of horizontal multipaths and ray chaos due to ocean mesoscale structure,'' J. Acoust. Soc. Am. **107**, 154–162 (2000).
[19] A. L. Virovlyansky and G. M. Zaslavsky, ''Evaluation of the smoothed interference pattern under conditions of ray chaos,'' Chaos **10**, 211–223 (2000).
[20] M. A. Wolfson and S. Tomsovic, ''On the stability of long-range sound propagation through a structured ocean,'' J. Acoust. Soc. Am. **109**, 2693–2703 (2001).
[21] I. P. Smirnov, A. L. Virovlyansky, and G. M. Zaslavsky, ''Theory and application of ray chaos to the underwater acoustics,'' Phys. Rev. E **64**, 366221 (2001).
[22] M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, ''Ray dynamics in long-range deep ocean sound propagation,'' J. Acoust. Soc. Am. **113**, 2533–2547 (2003).
[23] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, B. M. Howe, J. A. Mercer, R. C. Spindel, J. A. Colosi, K. Metzger, T. Birdsall, and A. B. Baggeroer, ''A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[24] J. A. Colosi, E. K. Scheer, S. M. Flatté, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. Birdsall, and A. B. Baggeroer, ''Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3202–3218 (1999).
[25] J. A. Colosi, F. Tappert, and M. Dzieciuch, ''Further analysis of intensity fluctuations from a 3252-km acoustic propagation experiment in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **110**, 163–169 (2000).
[26] J. L. Spiesberger, R. C. Spindel, and K. Metzger, ''Stability and identification of ocean acoustic multipaths,'' J. Acoust. Soc. Am. **67**, 2011–2917 (1980).
[27] P. F. Worcester, B. D. Cornuelle, J. A. Hildebrand, W. S. Hodgkiss, Jr., T. F. Duda, J. Boyd, B. M. Howe, J. A. Mercer, and R. C. Spindel, ''A comparison of measured and predicted broadband acoustic arrival patterns in travel time-depth coordinates at 1000 km range,'' J. Acoust. Soc. Am. **95**, 3118–3128 (1994).
[28] T. F. Duda, S. M. Flatté, J. A. Colosi, B. D. Cornuelle, J. A. Hildebrand, W. S. Hodgkiss, Jr., P. F. Worcester, B. M. Howe, J. A. Mercer, and R. C. Spindel, ''Measured wavefront fluctuations in 1000-km pulse propagation in the Pacific Ocean,'' J. Acoust. Soc. Am. **92**, 939–955 (1992).
[29] J. A. Colosi, S. M. Flatté, and C. Bracher, ''Internal-wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison with experiment,'' J. Acoust. Soc. Am. **96**, 452–468 (1994).
[30] M. A. Wolfson and J. L. Spiesberger, ''Full wave simulation of the forward scattering of sound in a structured ocean: A comparison with observations,'' J. Acoust. Soc. Am. **106**, 1293–1306 (1999).
[31] S. Flatté, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *Sound Transmission through a Fluctuating Ocean* (Cambridge University Press, Cambridge, 1979).
[32] J. A. Colosi and S. M. Flatté, ''Mode coupling by internal waves for multimegameter acoustic propagation in the ocean,'' J. Acoust. Soc. Am. **100**, 3607–3620 (1996).
[33] V. Cerveny, *Seismic Ray Theory* (Cambridge University Press, Cambridge, 2001).
[34] W. H. Munk, ''Internal waves and small scale processes,'' in *Evolution of Physical Oceanography*, edited by B. A. Warren and C. Wunsch (MIT Press, Cambridge, 1981), pp. 264–291.
[35] J. A. Colosi and M. G. Brown, ''Efficient numerical simulation of stochastic internal-wave-induced sound-speed perturbation fields,'' J. Acoust. Soc. Am. **103**, 2232–2235 (1998).
[36] J. G. Watson, W. L. Siegman, and M. J. Jacobsen, ''Refracted ray characteristics from the geometrical spreading equation for depth-dependent media,'' J. Acoust. Soc. Am. **70**, 583–590 (1981).
[37] W. H. Munk, ''Sound channel in an exponentially stratified ocean with application to SOFAR,'' J. Acoust. Soc. Am. **55**, 220–226 (1974).
[38] A. L. Virovlyansky, ''Ray travel times at long ranges in acoustic waveguides,'' J. Acoust. Soc. Am. **113**, 2523–2532 (2003).
[39] A. L. Virovlyansky, ''Description of stochastic ray motion in terms of action-angle variables,'' J. Acoust. Soc. Am. (submitted).
[40] J. A. Colosi, ''A review of recent results on ocean acoustic wave propagation in random media: Basin scales,'' IEEE J. Ocean. Eng. **24**, 138–155 (1999).
[41] M. V. Berry, ''Focusing and twinkling: Critical exponents from catastrophes in non-Gaussian random short waves,'' J. Phys. A **10**, 2061–2081 (1977).
[42] J. G. Walker, M. V. Berry, and C. Upstill, ''Measurements of twinkling exponents of light focused by randomly rippling water,'' Opt. Acta **30**, 1001–1010 (1983).
[43] J. H. Hannay, ''Intensity fluctuations beyond a one-dimensional random refracting screen in the short-wavelength limit,'' Opt. Acta **29**, 1631–1649 (1982).
[44] F. J. Beron-Vera and M. G. Brown, ''Ray stability in weakly range-dependent sound channels,'' J. Acoust. Soc. Am. **114**, 123–130 (2003).
[45] N. R. Cerruti and S. Tomsovic, ''Sensitivity of wave field evolution and manifold stability in perturbed chaotic systems,'' Phys. Rev. Lett. **88**, 054103 (2002).

Experimental and numerical study of air-coupled surface waves generated above strips of finite impedance[a)]

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

(Received 4 October 2002; revised 18 May 2003; accepted 16 June 2003)

A series of laboratory experiments are described in which air coupled surface waves are generated from a point source in the frequency range between 800 and 1700 Hz above a surface composed of a lattice of small cavities. Since the sound pressure near the lattice of cavities can be greater than if the surface was rigid, passive amplification is obtained. Moreover, directional receivers can be designed by restricting the lattice of cavities to a strip of finite width. These directional receivers are shown to provide more amplification than would be obtained with a semi-infinite lattice of cavities. Theoretical calculations using a boundary element method predict gains of up to 6 dB, relative to the pressure on a rigid plane when the strip width is of the order of 1 wavelength. The strips are further modified by changing their shape or by adding reflectors to further enhance the amplification. Gains of 12 dB have been measured and predictions show that gains of 20 dB are theoretically possible. [DOI: 10.1121/1.1603764]

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

I. INTRODUCTION

A surface composed of a lattice of small cavities can support the propagation of air-coupled surface waves. Their properties, and the applicability of theoretical models that describe them, have recently been summarized[1] and further studied[2,3] (see Ref. 1 for a more complete list of references to papers discussing air-coupled surface waves). In particular, these papers show that energy is localized near the surface and, since the sound pressure near the surface can be greater than if the surface was rigid, passive amplification is obtained. If the lattice of cavities is restricted to a strip of finite width embedded in an otherwise rigid surface, amplification will depend on the direction of the incident sound.[4] Thus, directional receivers can be designed that provide an alternative to beamforming techniques for long-range detection of acoustic signatures. These directional receivers can also provide more amplification than would be obtained with a semi-infinite lattice of cavities. The enhanced amplification also provides an improved signal-to-noise ratio.

In this paper, we discuss various ways in which strips of finite impedance can be configured to enhance the amplification (and directivity) properties of surface waves. Experiments were carried out on a model surface and the measurements are compared with numerical results obtained using a boundary element method.

II. THEORY AND NUMERICAL CALCULATIONS

Numerical calculations were performed using a boundary element method (BEM). A commercial software package SYSNOISE (version 5.5) was used for this purpose, running on a personal computer. The study of surfaces with regions of different impedance is not a typical application for this software, particularly with the presence of surface waves that propagate slower than the free-space speed of sound. A series of test calculations was performed to establish the applicability and accuracy of the computational software. First, a uniform impedance plane with a single point source, as sketched in Fig. 1(a), was considered. Numerical solutions can be compared directly to both exact and approximate theoretical solutions. Second, the semi-infinite impedance plane sketched in Fig. 1(b) was considered with a point source located above the side having zero admittance. This geometry has approximate theoretical solutions that can be used for comparison. Once the software is validated for these two geometries, it can be applied with confidence to the real situation of interest, the strip of finite impedance shown in Fig. 1(c).

An infinite impedance plane cannot be modeled directly with the numerical software. Instead, regions of finite extent are defined, meshed, and assigned impedances, and the surrounding plane modeled as acoustically rigid. Examples of such finite regions are indicated in Figs. 1(a) and (b) by the dashed lines, for infinite and semi-infinite cases. The impedance regions are increased in size and sound field calculations repeated, until the computed sound pressures converge. For the values of impedance used in this study, finite size (edge) effects were found to be negligible provided the size of the finite region was about 20 times larger than the incident wavelength, for field points not close to the edges. It might be noted that this approach to modeling an infinite impedance plane may not work if the resistive component of the impedance is too small—in such a case, edge-diffracted surface waves will propagate back toward the origin unattenuated.

There are several modules in SYSNOISE that permit the kinds of calculations that will be described here. It was de-

[a)]Portions of this work were presented in "The measurement of air-coupled surface waves over a porous surface," Proceedings of INTER-NOISE 93, Leuven, Belgium.

[b)]Now at the School of Information Technology and Engineering, University of Ottawa, Ottawa, ON.
[c)]Author to whom correspondence should be addressed. Electronic mail: gilles.daigle@nrc.ca

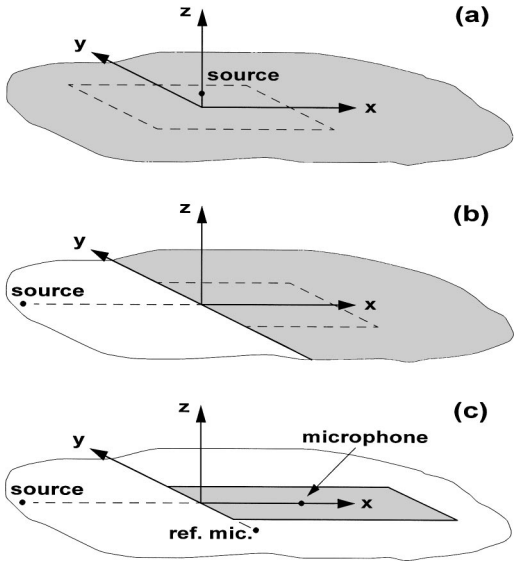


FIG. 1. Sketch of the geometries considered to validate the BEM calculations and used for the experiments. (a) A uniform impedance plane with a single point source located at $z_s=0.02$ m. (b) A semi-infinite impedance plane with a point source located $x_s=-4$ m and $z_s=0.5$ cm above the side having zero admittance. (c) A strip of finite impedance used for experiments. The point source is located at $x_s=-4$ m and $z_s=0.5$ cm. The reference microphone is located on the hard surface adjacent to the impedance discontinuity. The height of the measurement microphone is $z=0.5$ cm.

termined that the best compromise between speed of computation, flexibility and accuracy was achieved using the ''direct (collocation)'' BEM option. To implement this module, a planar mesh of elements in the $z=0$ plane (ground surface) is used to represent the rectangular region of interest and a plane of symmetry invoked in the $z=0$ plane to effectively place the mesh in an infinite rigid baffle.

To ensure accuracy in the calculations, mesh elements had linear dimensions less than 1/8 the wavelength of sound (free field) for the highest frequencies of interest—typically 2 cm. As a result, the number of elements could be very large and computation times could be very long. Where possible, use was made of symmetry to reduce the computational load. For the geometry of Fig. 1(a), symmetry planes about $x=0$ and $y=0$ were exploited. For Fig. 1(b), symmetry about $y=0$ was used.

The real and imaginary parts of the specific surface impedance, $Z/\rho c$, where ρc is the characteristic impedance of air, are listed in Table I for the frequencies considered in the calculations (the model surface for which this impedance is appropriate will be discussed later). Table I also gives the corresponding wavelength of the surface wave (see Ref. 1).

TABLE I. Normalized impedance used in the calculations and corresponding wavelength of the surface wave.

Frequency (Hz)	$Z/\rho c$	Wavelength (cm)
800	$0.651+4.48i$	41.9
1000	$0.543+3.59i$	33.1
1200	$0.463+2.96i$	27.2
1500	$0.378+2.30i$	21.1
1700	$0.338+1.96i$	18.1
1800	$0.321+1.82i$	16.9

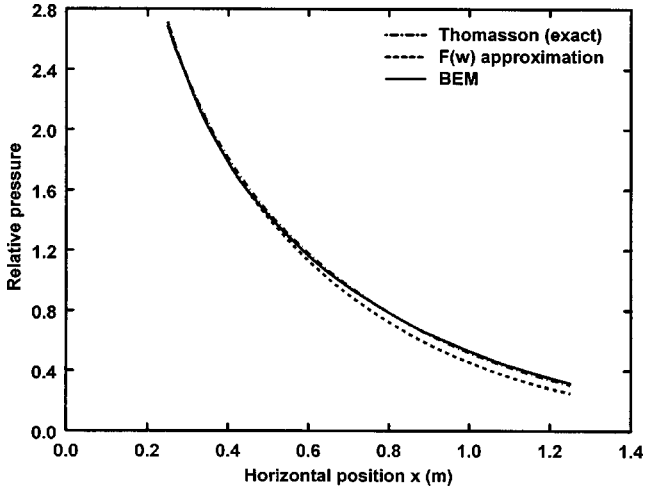


FIG. 2. Comparison between a BEM calculation and two analytical solutions in the case of a uniform impedance plane for a frequency of 1800 Hz. The point source is located at $z_s=0.02$ m. The longitudinal pressure variation is shown at a height $z=0.5$ cm and is relative to the free-field pressure. The curve labeled Thomasson (exact) is Eq. (5) and $F(w)$ approximation is Eq. (1).

A. Impedance plane

Referring to Fig. 1(a), we consider an infinite impedance plane at $z=0$, with a point source located a height z_s above the origin. Air-coupled surface waves are more strongly generated by keeping the source close to the impedance plane; heights as small as 2 cm were used in the validations. Tests were run for several frequencies between 800 Hz and 1800 Hz. Sound pressures were determined for a grid of field points in the *x-z* plane, the lower points being 0.5 cm above the surface.

Figure 2 shows a typical comparison between the BEM calculation and two analytical solutions. The solid curve is the result of a BEM calculation at 1800 Hz for a 4.5 m $\times$4.5 m impedance surface area with a point source located at $z_s=0.02$ m. The longitudinal pressure variation at a height of 0.5 cm is shown. The relative pressure has been obtained by dividing the calculated pressure by the sound pressure that would have been obtained if source and receiver only were present in free field. Surface waves can propagate over a surface with this impedance;[1] moreover, with the low source height, the surface wave component of the total pressure is significant.

The BEM calculation was first compared to a widely-used approximation for propagation from a point source above an impedance plane. The velocity potential is [see Fig. 3(a) for geometrical definitions]

$$\psi=\frac{e^{ikR_1}}{4\pi R_1}+Q\frac{e^{ikR_2}}{4\pi R_2}, \tag{1}$$

where k is the wave number and

$$Q=R_p+(1-R_p)F(w). \tag{2}$$

In Eq. (2), $R_p=(\cos\theta-\rho c/Z)/(\cos\theta+\rho c/Z)$ is the plane wave reflection coefficient and the boundary loss factor $F(w)$ is given by

$$F(w)=1+i\pi^{1/2}w\exp(-w^2)\mathrm{erfc}(-iw), \tag{3}$$

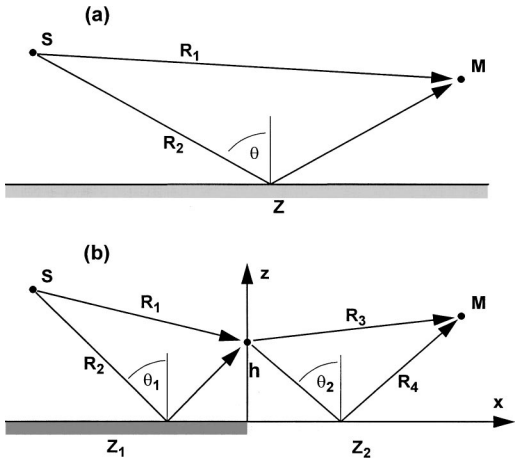


FIG. 3. Geometrical definitions. (a) Uniform impedance plane. (b) Approach developed by Rasmussen to calculate the sound field across an impedance jump.

where

$$w=(ikR_2/2)^{1/2}(\cos\theta+\rho c/Z) \tag{4}$$

is the numerical distance and erfc$(-iw)$ is the complementary error function (see Ref. 5 for computational details). The result from Eq. (1) for this case is shown by the dashed curve in Fig. 2. The agreement is good but there is a small deviation after a few wavelengths of propagation.

Thomasson[6] gives an exact integral solution for the field from a point source above an impedance plane. The velocity potential is

$$\psi=\frac{e^{ikR_1}}{4\pi R_1}+\frac{e^{ikR_2}}{4\pi R_2}+\Phi_d+\Phi_s. \tag{5}$$

In Eq. (5) the first term is the field from the source in the absence of the impedance plane, the second term represents the field reflected from a perfectly reflecting plane, the third term represents the field diffracted due to the finite impedance of the plane and is given by

$$\Phi_d=\frac{k\beta}{2\pi}I_d\exp(ikR_2), \tag{6}$$

where $\beta=1/Z$ and

$$I_d=\int_0^\infty\frac{e^{-t}}{W^{1/2}(t)}dt \tag{7}$$

is an integral that must be numerically evaluated. In Eq. (7)

$$W^{1/2}(t)=[(A^2+t)(B^2-t)]^{1/2} \tag{8}$$

where $A=[ikR_2(\gamma_0-1)]^{1/2}$ and $B=[ikR_2(1-\gamma_1)]^{1/2}$ with

$$\{\gamma,\gamma_1\}=-\beta\cos\theta\pm(1-\beta^2)^{1/2}\sin\theta \tag{9}$$

(see Ref. 7 to determine the choice of root $W^{1/2}$).

The fourth term in Eq. (5) represents the possible contribution from a surface wave:

$$\Phi_s=\tfrac{1}{4}(1-C)k\beta H_0(kx(1-\beta^2)^{1/2}]\exp(-ik\beta(z_s+z_r). \tag{10}$$

The variable $C=\pm1$ depending on whether $-\pi/2\leqslant\mathrm{Arg}(A)\leqslant\pi/4$ or $\pi/4\leqslant\mathrm{Arg}(A)\leqslant\pi/2$.

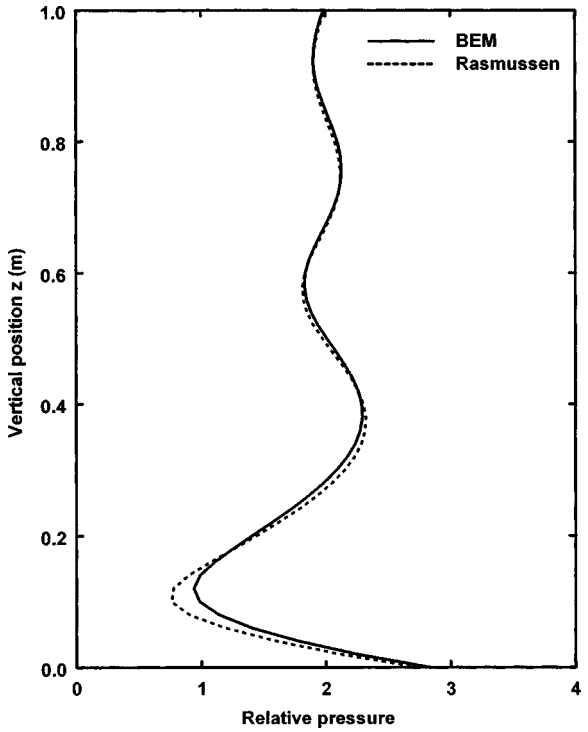


FIG. 4. Comparison between a BEM calculation and the theory of Rasmussen [Eq. (11)] for a frequency of 1500 Hz in the case of an impedance jump. The source is located 0.5 cm above the hard surface and 4 m from the discontinuity. The calculated pressure is sampled along a vertical line located at $x=0.5$ m and is relative to the free-field pressure.

The result from this exact theory, Eq. (5), is shown by the dashed-dot curve in Fig. 2. The BEM result almost overlaps with the exact result from Thomasson, demonstrating that the BEM software can accurately deal with an impedance plane, particularly one that supports the propagation of surface waves. The calculation was also done using the 2D axi-symmetric BEM module, taking advantage of the circular symmetry of the problem. With a 16 m diameter circular impedance area, essentially the same results are obtained. This module is computationally much faster than the 3D BEM calculation but it will not be applicable to the other geometries considered in this paper. The result shown in Fig. 2 was also obtained using ''indirect (variational)'' option of the software.

B. Impedance discontinuity

The second validation case treats the impedance discontinuity shown in Fig. 1(b). The $z=0$ plane has a surface impedance Z_2 for $x\geqslant0$ and is acoustically rigid ($Z_1=\infty$) for $x<0$. The solid curve in Fig. 4 shows the result for the BEM calculation for a frequency of 1500 Hz in the case of an impedance discontinuity [Fig. 1(b)]. The impedance area simulating the semi-infinite impedance surface measures 4.5 m$\times$4.5 m. The source is located 0.5 cm above the hard surface and 4 m from the discontinuity. The calculated pressure is sampled along a vertical line located 0.5 m away from the discontinuity and along the central line of the rectangular impedance area. The relative pressure is obtained by dividing the calculated pressure by the free-space pressure. The formation of a surface wave near the surface is clearly shown

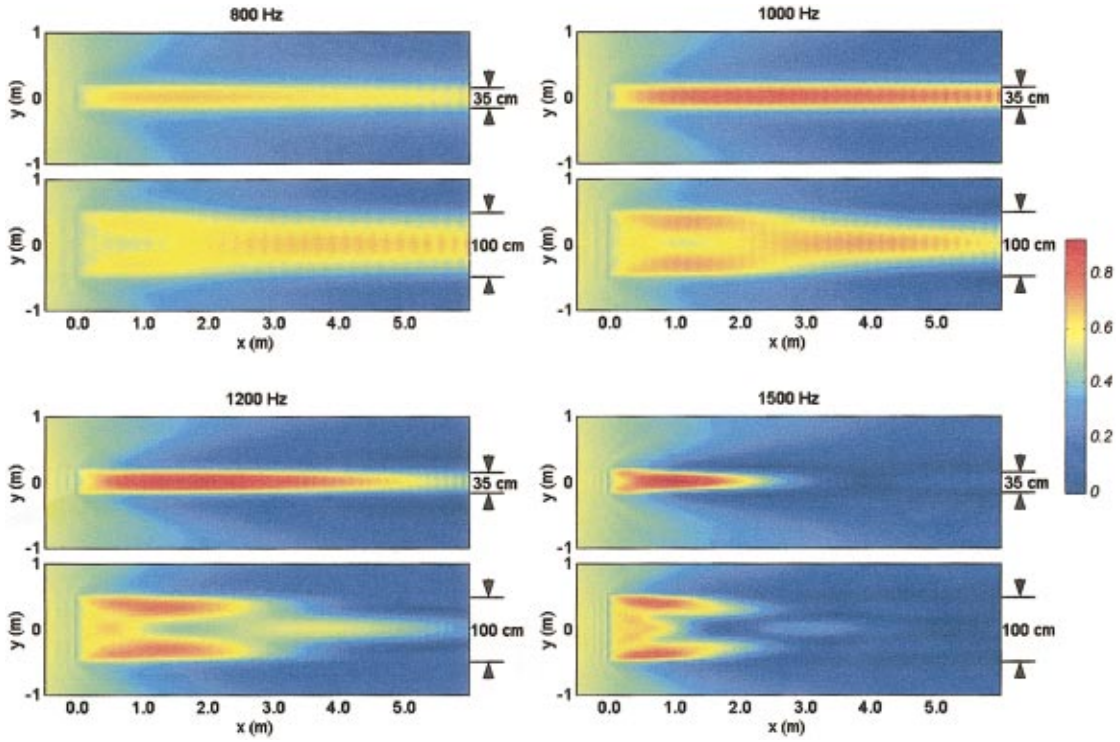


FIG. 5. Color plots of the pressure (magnitude, uncalibrated) distribution in the case of four frequencies and two impedance strip widths (indicated at the right of each panel). The strips begin at $x=0$ and the pressure is sampled at a height $z=0.5$ cm. The source is located 0.5 cm above the hard surface and 4 m to the left of the strip along its central line ($x_s=-4$ m and $y_s=0$).

(see Ref. 2 for a fuller discussion of the characteristics of the curve and surface wave formation at a discontinuity).

Rasmussen[8] provides an approximate theoretical formulation for propagation of sound over an impedance jump. An intermediate vertical plane is introduced at the discontinuity as shown in Fig. 3(b), and for each position h on this plane the four paths R_m are constructed ($m=1-4$). The velocity potential is

$$\psi = x(8\pi k)^{1/2}\frac{e^{-i\pi/4}}{16\pi^2}\int_0^\infty\left[\frac{e^{ik(R_1+R_2)}}{(R_3^3R_1(R_1+R_3))^{1/2}} + Q_2\frac{e^{ik(R_1+R_4)}}{(R_3^3R_1(R_1+R_3))^{1/2}} + Q_1\frac{e^{ik(R_2+R_3)}}{(R_3^3R_2(R_2+R_3))^{1/2}} + Q_1Q_2\frac{e^{ik(R_2+R_4)}}{(R_4^3R_2(R_2+R_4))^{1/2}}\right]dh. \quad (11)$$

In Eq. (11) $Q_{1,2}$ is calculated using Eq. (2) for both surfaces. Guidelines for the numerical integration of Eq. (11) are discussed by Rasmussen.[8] The dashed curve in Fig. 4 was obtained from Eq. (11). The agreement between the BEM and Eq. (11) is good. With these tests, those in Subsection A, and others, we feel confident in applying the numerical package to strips.

C. Strips of finite width

Our main concern in this paper is impedance strips of finite length and width. Their amplification properties are demonstrated in this section. Figure 5 gives a qualitative picture of the effect of sound frequency and strip width. The figure shows color plots of the BEM simulated pressure (magnitude) distribution in the $x-y$ plane at a height of $z=0.5$ cm. Four frequencies and two strip widths are considered. The strips begin at $x=0$ cm and the widths are 35 and 100 cm as indicated to the right of each color plot. The source is located 0.5 cm above the hard surface and 4 m to the left of the strip along its central line, i.e., $x_s=-4$ m and $y_s=0$. The regions in red are areas of highest sound pressure (uncalibrated) while regions in blue are areas of lowest sound pressure, as indicated by the color bar to the right of the figure.

The lateral confinement of acoustic energy into a propagating surface wave is evident in the different panels. The amplification shown in Fig. 5 is also seen here by the regions of red at 1000, 1200, and 1500 Hz (wavelengths of 33.1, 27.2, and 21.1 cm, see Table I) along the strip in the case of the 35 cm width. When the strip width is increased to 100 cm, the pressure distribution along the strip has a more complex modal structure. For example, at 1200 Hz there is amplification in a localized region at the beginning of the strip near the two edges to the rigid surface and there is a dip in pressure in the center of the strip around $x=2$ m. The impedance of the strip has not been optimized for surface wave formation at 800 Hz. For all the plots, the attenuation of the surface wave at larger x, due to the resistive component of the surface impedance, can be seen. The plots also show interference patterns: there is a surface wave reflected from the end of the strip. The attenuation is sufficient that little reflected wave is found in the first 2 m of the strips where more detailed results are obtained.

The qualitative results illustrated in Fig. 5 are quantified in Fig. 6. To start, a baseline prediction for a semi-infinite

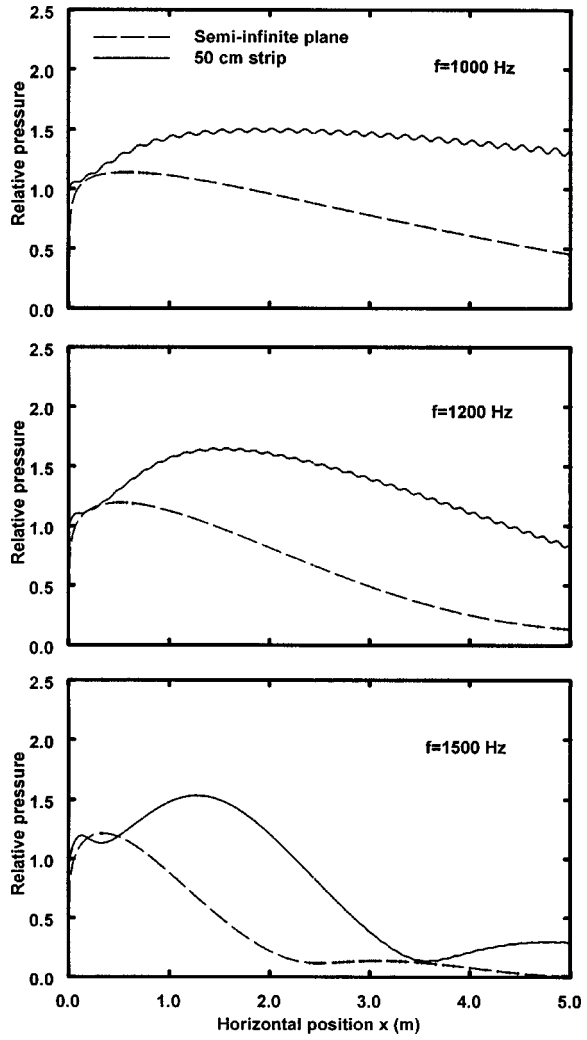


FIG. 6. Comparison between BEM calculations in the case of a 50 cm wide strip and the theory of Rasmussen [Eq. (11)] for a semi-infinite impedance plane, showing the longitudinal variation. The source is located at $x_s=-4$ m at a height of $z_s=0.5$ cm and the receiver height is $z=0.5$ cm. The pressure is relative to the pressure calculated at a distance of 4 m from a source above a hard surface.

impedance is obtained using the theory of Rasmussen: the dashed curves in Fig. 6 were calculated using Eq. (11) for frequencies of 1000, 1200, and 1500 Hz. The source is located at $x_s=-4$ m at a height $z_s=0.5$ cm above the hard surface. The receiver height is 0.5 cm. The pressure plotted is relative to the pressure calculated at a distance of 4 m from a source above a hard surface (i.e., at the beginning of the discontinuity, but in the absence of the strip). This reference pressure was chosen to show that the surface wave initially grows with increasing horizontal position ($x>0$), reaching pressures at positions x between 20 and 50 cm that exceed the pressures calculated at $x=0$ cm above the hard surface. Thus passive amplification is obtained with a maximum gain in pressure of about 1.2 [or $20\log(1.2)\approx 1.6$ dB] at 1200 Hz in the case of this impedance.

The solid curves in Fig. 6 illustrate the enhanced amplification possible by using strips of finite width. They were obtained from the BEM by assuming an impedance strip of width 50 cm [see Fig. 1(c)]. The increased amplification at all three frequencies is clearly evident with the largest gain in

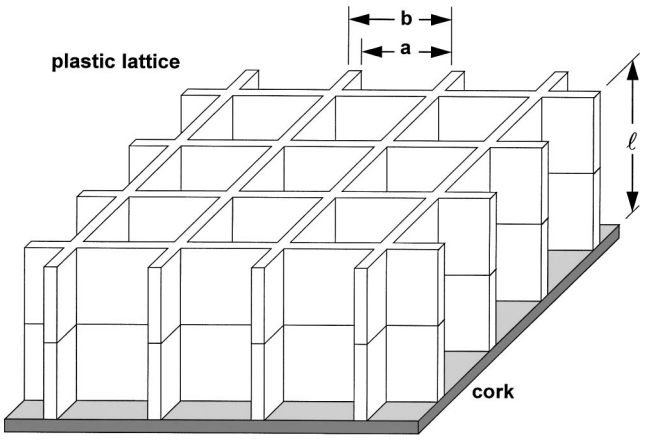


FIG. 7. Sketch showing the construction of the model surface. The relevant dimensions are $a=1.315$ cm, $b=1.5$ cm, and $\ell=2.56$ cm.

pressure of about 1.6 (or 4.1 dB) at 1200 Hz. The small oscillations, especially for 1000 and 1200 Hz, are attributed to interference of the surface wave with a wave reflected from the end of the strip.

By restricting the lattice of cavities to a long but narrow strip of finite width embedded in an otherwise rigid surface, the amplification will depend on the direction of the incident sound [see Fig. 1(c)]. If the strip is sufficiently narrow to only allow space for surface wave formation in the direction of the strip length, the strip becomes a highly directional receiver. Since the phase speed of the surface wave along the strip is less than the speed of sound above the adjacent rigid surfaces each side of the strip, the sound energy each side of the strip is expected to contribute to the surface wave formation and thus enhance the amplification. The directional characteristics of these strips will be presented experimentally later in this paper.

III. MEASUREMENT PROCEDURE

The model surface used in the experiments was constructed from commercial overhead lighting panels as indicated in Fig. 7. Two panels are placed one on top of the other and fastened with screws to a thin layer of cork board glued to plywood sheets. The relevant dimensions are $a=1.315$ cm, $b=1.5$ cm, and $\ell=2.56$ cm. Full details of the construction have been described previously in Ref. 1 and the effective surface impedance has been shown to be

$$Z=i\rho c\cot\left(\frac{k\ell}{2}\right)\frac{1-q\tan^2(k\ell/2)-\varepsilon\tan(k\ell/2)\sin^2\theta}{1+q+\varepsilon\cot(k\ell/2)\sin^2\theta}\left(\frac{b}{a}\right)^2, \tag{12}$$

where

$$q=\frac{1+i\rho c\cot(k\ell/2)/Z_s}{1-i\rho c\cot(k\ell/2)/Z_s} \tag{13}$$

accounts for the finite impedance of the underlying cork layer and

$$\varepsilon=\frac{i\rho c b^2}{a^2 Z_L}. \tag{14}$$

In Eq. (13), Z_s is the specific surface impedance of the cork. It is calculated using the single parameter empirical model of Delany and Bazley[9] by specifying[1] a flow resistivity of 1500 kPa·s/m^2 and the thickness of the cork layer (0.48 cm). In Eq. (14), Z_L accounts for the acoustical leak impedance between adjacent cells of the lattice. It is given by

$$Z_L = \frac{-i\omega\rho_s(\omega)(b-a+0.015)}{\eta b}, \tag{15}$$

where $\eta = 0.11$ cm is the dimension of the gap and

$$\rho_s(\omega) = \rho[1-\tanh(\sqrt{i}\lambda)/(\sqrt{i}\lambda)]^{-1} \tag{16}$$

with $\lambda = (\eta/2)(\omega\rho/\mu)^{1/2}$ where μ is the coefficient of viscosity for air [1.83×10^{-4} g/(m·s)]. Note that Eqs. (14) and (15) correct typographical errors found in the Appendix of Ref. 1. The impedance values used for the theoretical and BEM calculations in the previous Section were obtained from Eq. (12).

Measurements were performed in a large experimental space using gated pure tone pulses to remove the extraneous reflections. The source signal was fed to a compression driver fitted with a long tube to approximate an omnidirectional point source. The source is located on the ground at a position $x_s = -4$ m in front of the strip. The receiver was a $\frac{1}{2}$ -in. B&K microphone suspended above the surface. A reference microphone was placed on the hard surface adjacent to the impedance discontinuity [see Fig. 1(c)]. The position was chosen to minimize diffractive effects from the strip edges and the measured pressure is corrected (inverse square law) to obtain the pressure at a distance of 4 m from the source. Strips of various widths (between 100 and 10 cm) were constructed by overlaying large sheets of thin aluminum over the lighting panels, leaving exposed only a strip of cavities centered and along the positive x-axis. Measurements were made at frequencies between 400 and 1700 Hz.

IV. RESULTS FOR RECTANGULAR STRIPS

In this section, the propagation of bounded surface waves, presented in Fig. 5, is studied in more detail by looking at the pressure distributions along and across various strips. Unless otherwise indicated, the pressures are plotted relative to the pressure at a distance of 4 m from a source located above a hard surface. The source is located at $x_s = -4$ m and $z_s = 0.5$ cm and the height of the measurement microphone is 0.5 cm in all cases [see Fig. 1(c)].

A. Transverse variation

The data points in Fig. 8 were measured at a frequency of 1500 Hz (wavelength of 21.1 cm) by moving the microphone along the y-axis, for horizontal positions $x = 50$, 100, 150, and 200 cm. The strip width is 20 cm (the arrow at the bottom of the figure indicates the strip width). The measurements clearly show amplification (relative pressure exceeding unity) at the three shorter horizontal positions along the center ($y \approx 0$ cm) of the strip, with the largest amplification at 100 cm. As the transverse position of the microphone increases, the pressure drops rapidly to reach a minimum just beyond the edge of the strip. By a transverse position of 50 cm well outside of the strip, the measurements approach inverse square law with increasing x as expected above a rigid surface.

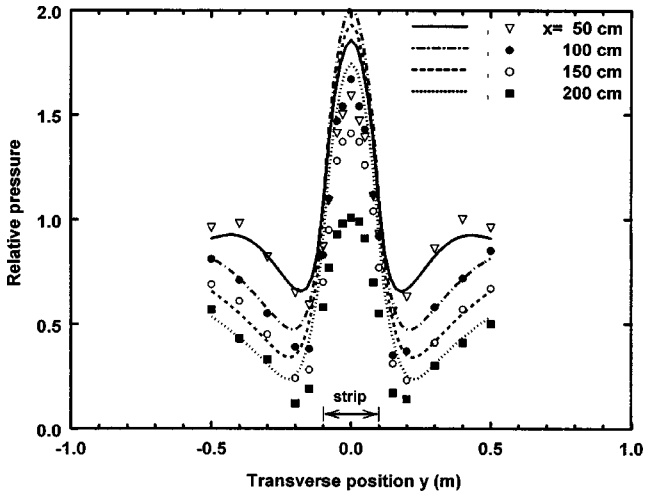


FIG. 8. The transverse pressure variation in the y-direction across a 20 cm wide strip for four horizontal positions. The frequency is 1500 Hz (wavelength of 21.1 cm). The source is located at $x_s = -4$ m at a height of $z_s = 0.5$ cm and the receiver height is $z = 0.5$ cm. The data points are the measured results and the curves are the corresponding BEM calculations. The pressure is relative to the pressure at a distance of 4 m from the source above a hard surface.

The curves in Fig. 8 are the corresponding BEM calculations. Although the predictions show the same tendencies as the measurements, the predicted amplification is larger, especially at $x = 200$ cm. Additional calculations suggests that the more rapid attenuation of the surface wave can be attributed to energy leaking underneath the aluminum sheets that are used to construct the strip. Both calculations and measurements show that, the narrower the strip, the more leakage is expected to affect the experiment. No attempt has been made to precisely quantify this artifact.

In Fig. 9, the data points show the measurements at four frequencies between 800 and 1500 Hz (wavelengths between 41.9 and 21.1 cm). This strip has a width of 50 cm and the results were obtained at a range of $x = 200$ cm. The measurements were only made in the positive y-direction and the results have been mirrored in the negative y-direction. The curves are the corresponding BEM calculations. The agreement between theory and measurement is good at all frequencies. The prediction still shows greater amplification, but the discrepancy is not as large as in the previous figure because of the larger strip width.

Figure 10 shows results for strip widths of 20, 50, and 100 cm. The data points are measurements and the curves are prediction. The frequency is 1500 Hz (wavelength of 21.1 cm) and the range is $x = 100$ cm. Note the complex pressure distribution across the larger strip widths, for both measurements and predictions. In the case of a 100 cm width, the pressure shows amplification close to the two edges and a pressure drop along the center of the strip (see Fig. 5). The wavelength for the surface wave at 1500 Hz is 21.1 cm, so a more complex modal structure in the pressure distribution is not unexpected.

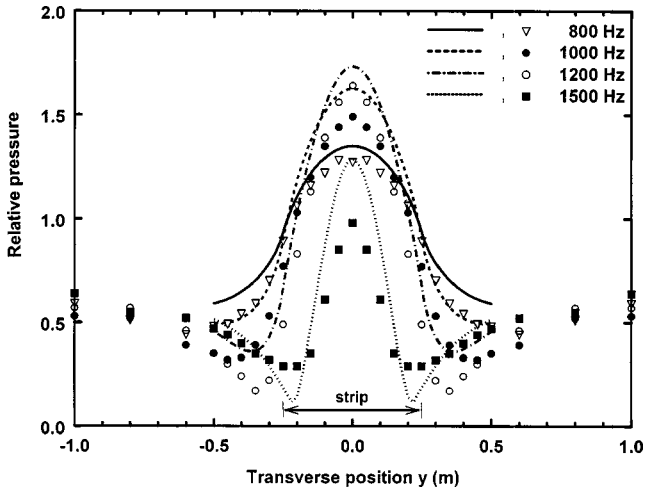


FIG. 9. The transverse pressure variation in the *y*-direction across a 50 cm wide strip in the case of four frequencies. The horizontal position is fixed at $x=2$ m. The source is located at $x_s=-4$ m at a height of $z_s=0.5$ cm and the receiver height is $z=0.5$ cm. The data points are the measured results and the curves are the corresponding BEM calculations. The pressure is relative to the pressure at a distance of 4 m from the source above a hard surface.

B. Longitudinal variation

In this subsection the variation of the surface wave amplification along the horizontal position x in the center ($y=0, z=0.01$ m) of the strip is studied. The data points in Fig. 11 are measurements for strips of width 20, 50, and 100 cm at frequencies between 800 and 1700 Hz (wavelengths between 41.9 and 18.1 cm). The curves are the corresponding BEM calculations. In all cases, the relative pressure is close to unity at $x=0$, rises as the surface forms, then reaches a local maximum before dropping. Good agreement between theory and measurements is generally obtained except in the case of the narrower strip, as noted in previous figures. The theoretical results suggest that a doubling of pressure (a gain of 6 dB) can be achieved at frequencies between about 1200 and 1500 Hz, at horizontal positions of about $x=1$ m for this impedance. By further reducing the resistive component of the impedance, more gain could be achieved. For an optimized surface with no resistance, an amplification of 7.4 dB is predicted.

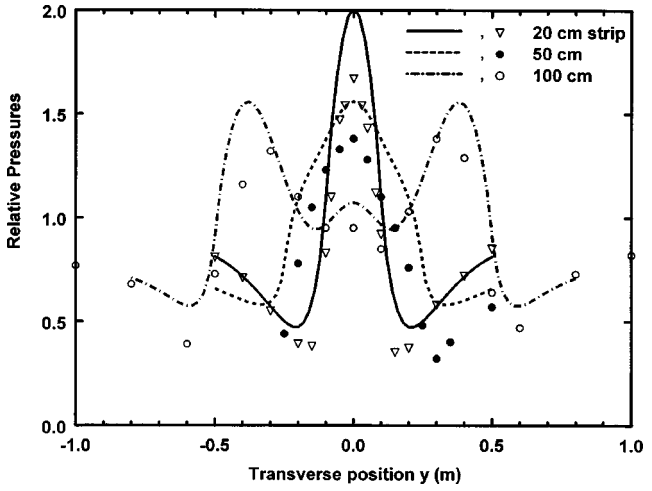


FIG. 10. The transverse pressure variation in the *y*-direction in the case of three strip widths for a horizontal position fixed at $x=1$ m and for a frequency of 1500 Hz (wavelength of 21.1 cm). The source is located at $x_s=-4$ m at a height of $z_s=0.5$ cm and the receiver height is $z=0.5$ cm. The data points are the measured results and the curves are the corresponding BEM calculations. The pressure is relative to the pressure at a distance of 4 m from the source above a hard surface.

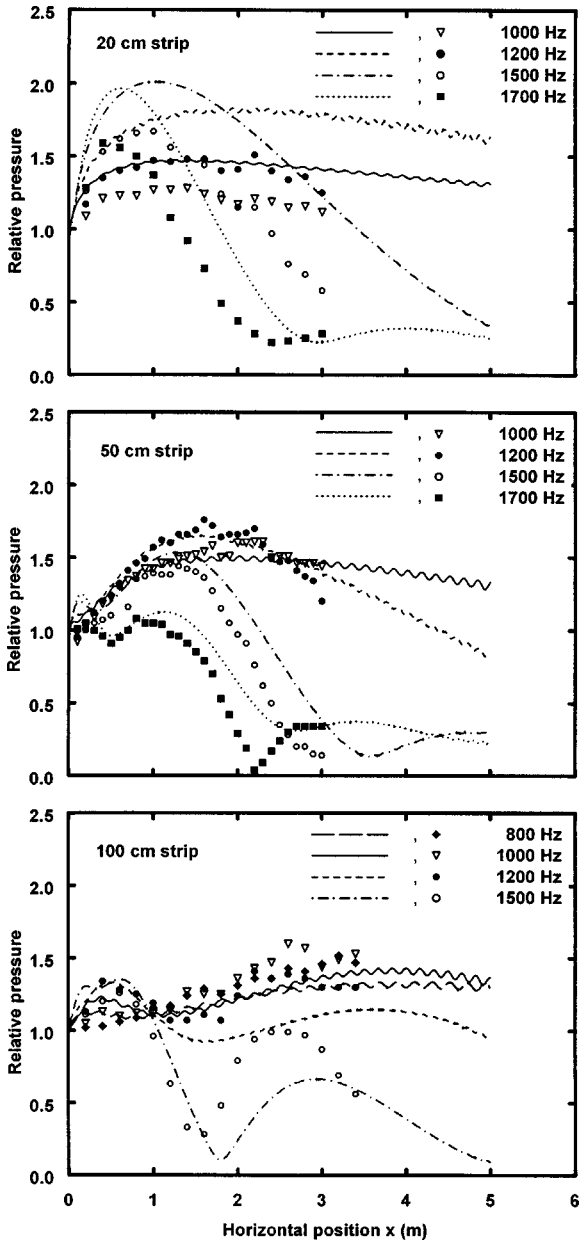


FIG. 11. The longitudinal pressure variation along the position x in the center ($y=0$) of the strips for widths of 20, 50, and 100 cm at frequencies between 800 and 1700 Hz (wavelengths between 41.9 and 18.1 cm). The source is located at $x_s=-4$ m at a height of $z_s=0.5$ cm and the receiver height is $z=0.5$ cm. The data points are the measured results and the curves are the corresponding BEM calculations. The pressure is relative to the pressure at a distance of 4 m from the source above a hard surface.

C. Effect of strip width

The results in Fig. 12 show the maximum surface wave amplification (measured and predicted) as a function of strip width for frequencies of 1000, 1200, and 1500 Hz (wavelengths of 33.1, 27.2, and 21.1 cm). The maximum amplification was obtained from Fig. 11 by finding the maximum

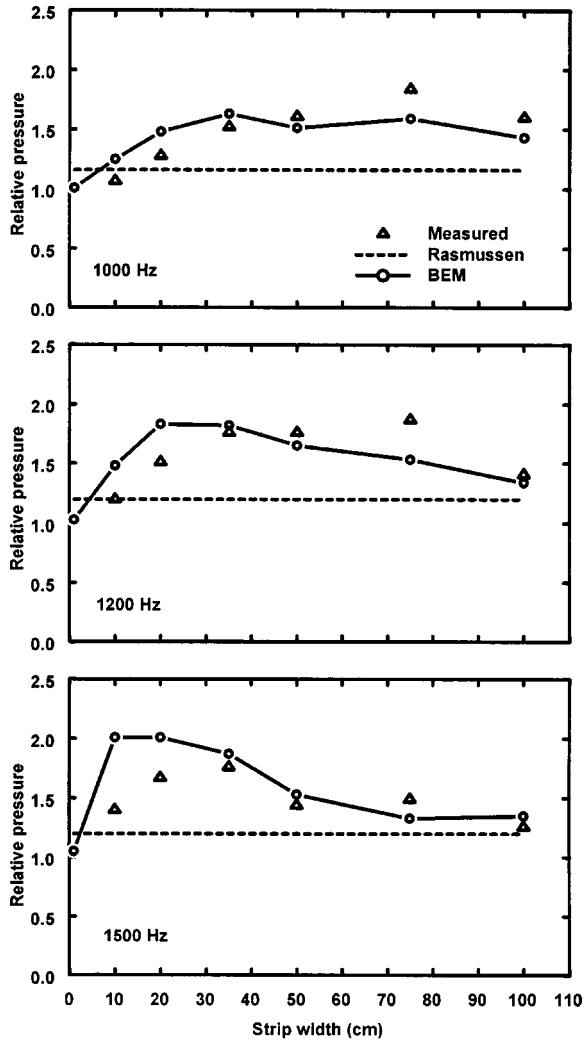


FIG. 12. The maximum surface wave amplification as function of strip width obtained by finding the maximum pressure at each frequency from Fig. 11. The dashed lines are the maximum relative pressures predicted from the theory of Rasmussen [Eq. (11)] in the case of a semi-infinite impedance surface.

pressure at each frequency for each strip width. For example, at 1200 Hz in the case of the 50 cm strip, theory and measurement in Fig. 11 show maximum relative pressures of about 1.6 and 1.7, respectively, at a horizontal position of about 1.5 m. In the case of the 20 cm strip, the theory predicts a maximum relative pressure of 2 at 1500 Hz at a horizontal position of about 1 m, while the measured maximum relative pressure is only about 1.6 at about the same distance.

The dashed lines in Fig. 12 are the maximum relative pressures that are expected in the case of a semi-infinite surface predicted using Rasmussen's theory Eq. (11), i.e., the maximum relative pressure taken from the dashed curves in Fig. 6 at distance between $x=0.2-0.5$ m. The results in Figs. 11 and 12 demonstrate that, for a given frequency, there is an optimum strip width where the greatest gain can be achieved at a given horizontal position, these width being about 15 cm at 1500 Hz, 25 cm at 1200 Hz, and 35 cm at 1000 Hz, i.e., when the strip width is about 1 wavelength. As the strip width increases beyond the optimum width the maximum pressure decreases and tends toward the value expected from Eq. (11) above a semi-infinite strip; as the width is decreased the maximum pressure approaches the value of unity corresponding to a hard surface,

D. Directivity

The measured directional characteristics of the strips are shown in Fig. 13. The data points were obtained for frequencies between 400 and 1500 Hz (wavelengths approximately between 85.7 and 21.1 cm) in the case of strips of width 53 and 86 cm. The source was rotated about the microphone position, in the *x-y* plane, keeping the source-microphone distance constant. The angle of 0° corresponds to the source pointing at the strip along the *x*-axis. The SPL is relative to the free field pressure. The solid line is drawn through the points. These results were obtained on an earlier surface for which the impedance was not known precisely and no attempt has been made to obtain BEM results.

The results in Fig. 13(b) for 1200 Hz show a sharp narrow beam pattern in the direction of the *x*-axis. Although not as sharp and strong, a beam pattern is also observed at 400 Hz in Figs. 13(a) and (c). The complex beam pattern measured at 1500 Hz in Fig. 13(d) is consistent with the results shown in Figs. 5 and 10.

V. OTHER STRIP GEOMETRIES AND MODIFICATIONS

In this section, the strips are modified in order to further enhance the amplification. The three cases studied are shown in Fig. 14 and include a tapered strip, the effect of adding a vertical plane across the strip, and the effect of adding a parabolic reflector across the strip.

A. Tapered strip

The tapered strip is shown in Fig. 14(a). The initial width at $x=0$ is 100 cm and the width is linearly tapered to zero at $x=3.51$ m. The data points in Fig. 15 were measured at 800 and 1500 Hz. The curves are the results of the BEM calculation. Agreement between measurements and calculations is fairly good. These results should be compared with those in Fig. 11 for the rectangular strip of width 100 cm. This is done explicitly in the case of the results at 800 Hz in Fig. 16. Both theory and measurement show addition amplification at horizontal positions between about 1.5 and 3 m in the case of the tapered strip, with a maximum additional gain of about $20\log(1.5)-20\log(1.26)\approx 1.5$ dB.

B. Vertical plane reflector

Figure 14(b) shows a plane reflector positioned at $x=2$ m perpendicular to a strip of width 100 cm. The sound source is at $x=-4$ m. The reflector was a sheet of plywood, 30 cm high and 120 cm long. For the BEM "direct" calculation, the reflector was modeled as one cell (2 cm) thick. In Fig. 17, the top panel (a) shows measurements, the bottom panel (b) shows the calculated results. The open circles in Fig. 17(a) were measured for various frequencies with the microphone positioned immediately in front of the reflector. The measurements were repeated in the absence of the reflector to obtain the open triangles. The vertical scale shows uncalibrated pressure so only the relative difference between the two sets of points in relevant in Fig. 17. The plane re-

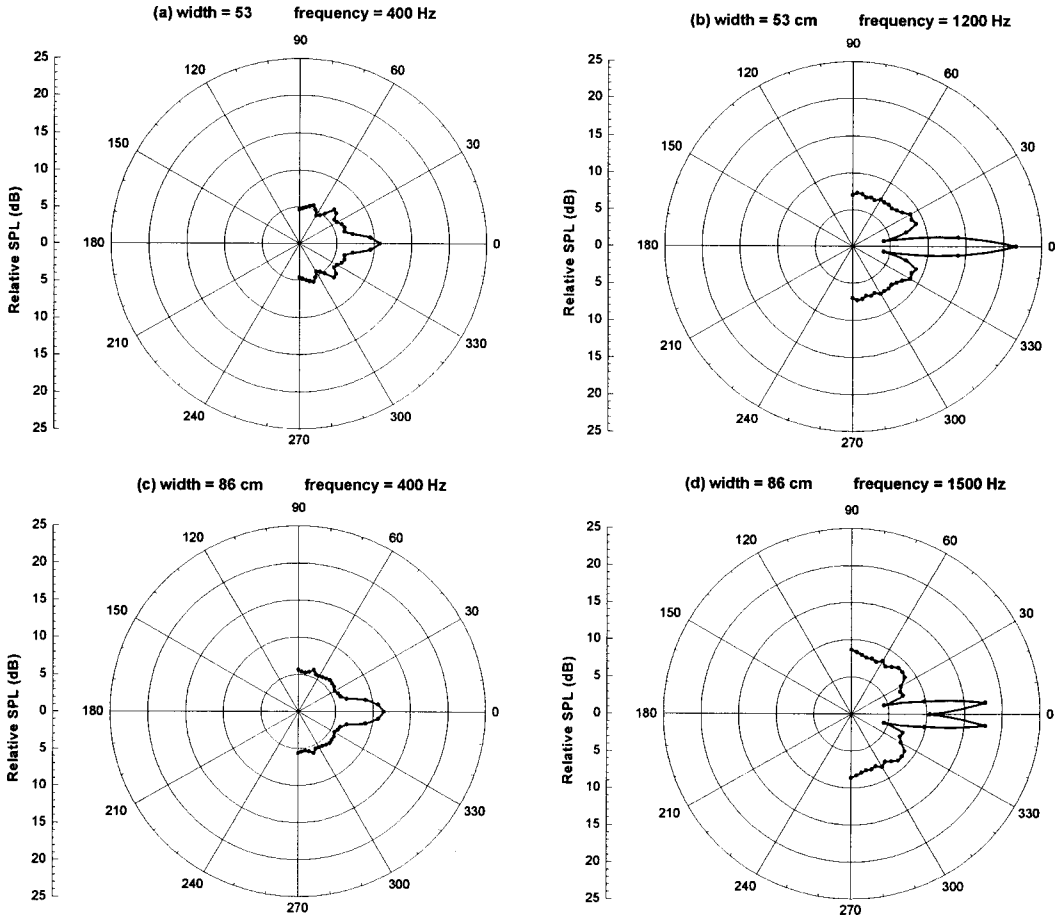


FIG. 13. The measured (points) directional characteristics of two strips for frequencies between 400 and 1500 Hz. The SPL is relative to the free field pressure and the circles are spaced every 5 dB. The solid line is drawn through the points.

flector provides additional gain with a maximum gain of about $20\log(2.5)-20\log(1.34)\approx 5.4$ dB measured at 920 Hz. The BEM calculations for the plane reflector are shown in Fig. 17(b). The calculated results show a slightly different frequency behavior, with the reflector providing a more uniform gain of about $20\log(1.23)-20\log(0.6)\approx 6.2$ dB at frequencies below 900 Hz.

These results provide clear evidence that surface waves can be reflected by a vertical structure. At the reflector, the sum of incident and reflected waves gives the observed ''pressure doubling.'' For the measured results, we believe that leakage under the reflector accounts for the somewhat reduced gain.

C. Parabolic reflector

The parabolic reflector is illustrated in Fig. 14(c). It was formed by bending a thin sheet of aluminum and using tape to hold it in this bent shape. Measurements confirmed that the shape was parabolic, with a focal distance of 20 cm. The overall width of the reflector was 100 cm and the depth of the bowing was 31.25 cm. The height was 41.4 cm. For the BEM ''direct'' calculations, the reflector was assumed to be one cell (1.8 cm) thick. In the experiment, the microphone was located 20 cm in front of the reflector, which is the focal point of the parabola. The open circles in Fig. 18(a) were measured at $x=200$ cm on a rectangular strip of 100 cm

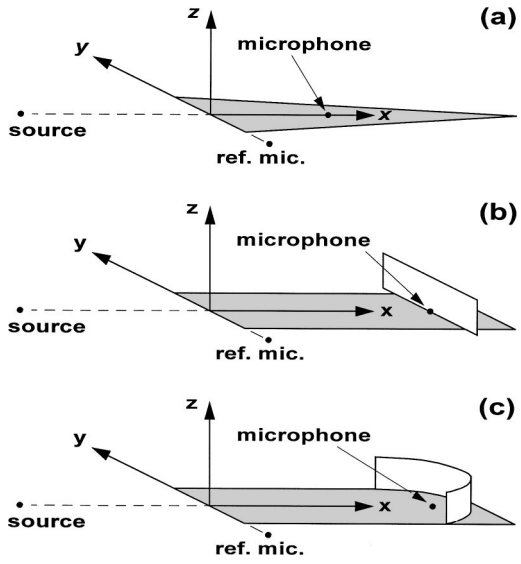


FIG. 14. Sketch of other strip geometries and modification. The source is located at $x_s=-4$ m and $z_s=0.5$ cm. The reference microphone is located on the hard surface adjacent to the strip. (a) A tapered strip. The initial width at $x=0$ is 100 cm and the width is linearly tapered to zero at $x=3.51$ m. (b) A plane reflector positioned at $x=2$ m above a strip of width 100 cm. The reflector was a sheet of plywood 30 cm high and 120 cm long. The microphone is positioned immediately in front of the reflector. (c) A parabolic reflector above a strip of width 100 cm. The reflector was formed by bending a thin sheet of aluminum. The overall width of the reflector was 100 cm and the depth of the bowing was 31.25 cm. The height was 41.4 cm. The microphone is located at the focal distance (20 cm) of the reflector.

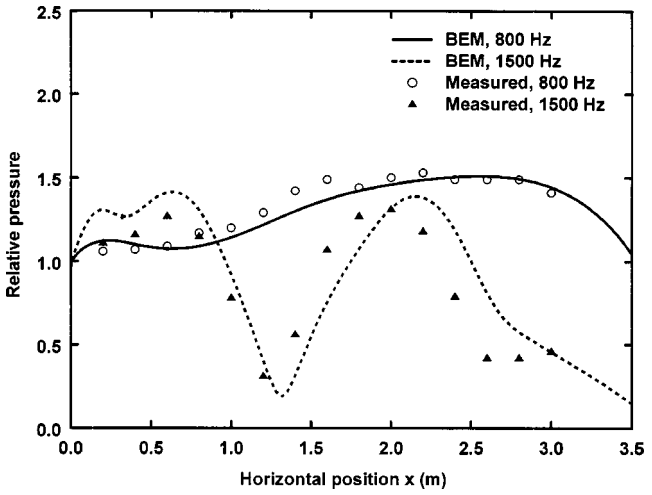


FIG. 15. Measurements and BEM calculations in the case of the tapered strip for frequencies of 800 and 1500 Hz (wavelengths of 41.9 and 21.1 cm). The pressure is relative to the pressure at a distance of 4 m from the source above a hard surface.

width for several frequencies with the reflector in place. The measurements were repeated without the reflector to yield the open triangles. An additional gain of $20\log(4) - 20\log(1.08) \approx 11.4$ dB is obtained for frequencies near 1000 Hz. The BEM calculations are shown in Fig. 18(b). The theory predicts an additional gain of about $20\log(2.25) - 20\log(0.54) \approx 12.4$ dB for frequencies near 1000 Hz.

Agreement between the measurements and calculations is quite good. The large gains obtained confirm that surface waves can be focused.

VI. CONCLUSION

Air-coupled surface waves were generated at an impedance discontinuity on a flat surface, one section of the surface being acoustically hard ($x<0$), and the other composed of a lattice of small cavities ($x>0$). When the surface waves are formed, the sound pressure above the lattice of cavities ($x>0$) can exceed the sound pressure at the discontinuity

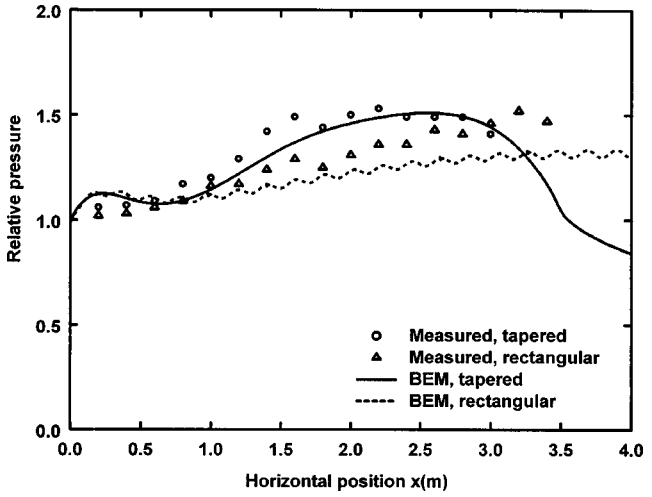


FIG. 16. Comparison between results obtained above the tapered strip and a rectangular strip of width 100 cm, in the case of 800 Hz. The pressure is relative to the pressure at a distance of 4 m from the source above a hard surface.

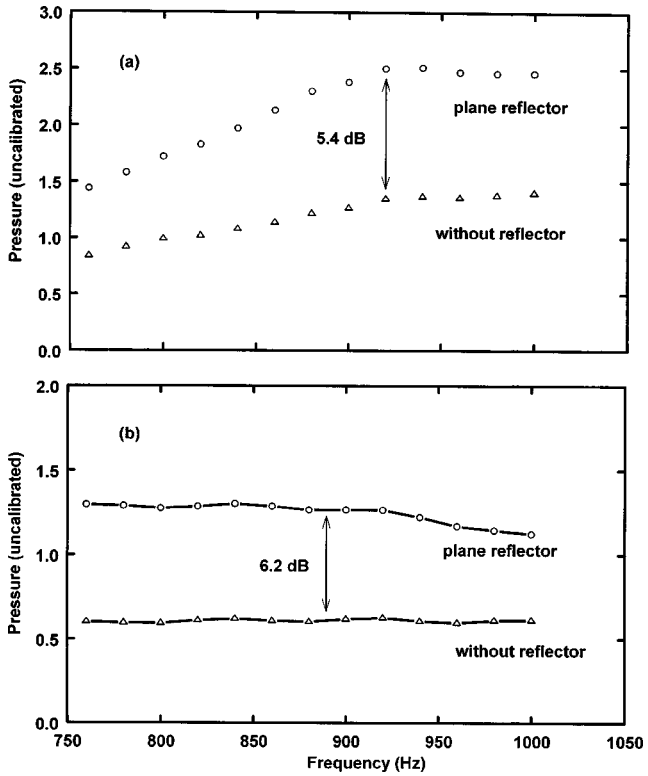


FIG. 17. Comparison between results obtained in the presence and absence of a plane reflector, for frequencies between 760 and 1000 Hz. (a) Measured results. (b) BEM calculations.

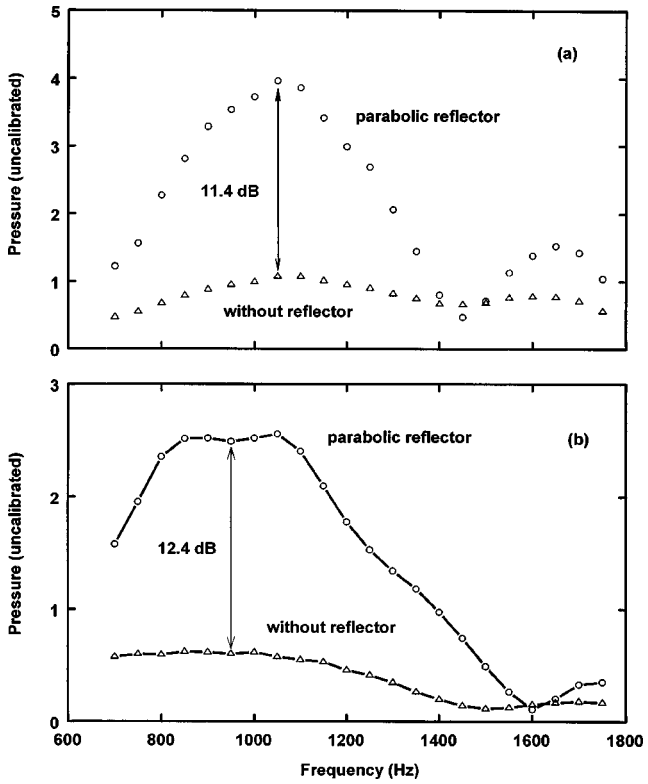


FIG. 18. Comparison between results obtained in the presence and absence of the parabolic reflector, for frequencies between 700 and 1750 Hz. (a) Measured results. (b) BEM calculations.

($x=0$) above the rigid surface and passive amplification is obtained. In the case of a semi-infinite lattice of cavities, a gain of about 1.6 dB is expected. By restricting the lattice of cavities to a strip of finite width embedded in an otherwise rigid surface, the amplification depends on the direction of the incident sound and directional receivers can be designed. The strips also provide more amplification than would be obtained with a semi-infinite lattice of cavities. By tuning the frequency and the width of the strip, gains of up to 5 dB have been measured while gains of up to 6 dB are theoretically possible. By further reducing the resistive component of the impedance, more gain could be achieved. For an optimized surface with no resistance, an amplification of 7.4 dB is predicted.

More amplification is possible by tapering the width of the strip or adding reflectors to focus the energy of the surface wave. The tapered strip gives an additional gain of 1.5 dB. The largest amplification has been obtained by using a parabolic reflector on a rectangular strip. On a strip of 100 cm width at a frequency of 1000 Hz, the measured signal is 12 dB greater than would be measured if the reflector were not present. By constructing a surface whose impedance has a smaller resistive component and optimizing strip width and parabolic shape, it is anticipated that an amplification of 20 dB can be achieved.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Tony Embleton for his suggestions regarding this work. Discussions with Dr. David Havelock are appreciated. We would like to thank René St. Denis for his efforts in constructing the model surface and for collecting data. This work was funded in part through a collaborative agreement with Defense Research Establishment Valcartier.

[1] G. A. Daigle, M. R. Stinson, and D. I. Havelock, ''Experiments on surface waves over a model impedance plane using acoustical pulses,'' J. Acoust. Soc. Am. **99**, 1993–2005 (1996). This paper contains a good list of references to papers on air-coupled surface waves.

[2] M. R. Stinson and G. A. Daigle, ''Surface wave formation at an impedance discontinuity,'' J. Acoust. Soc. Am. **102**, 3269–3275 (1997).

[3] W. Zhu, M. R. Stinson, and G. A. Daigle, ''Scattering from impedance gratings and surface wave formation,'' J. Acoust. Soc. Am. **111**, 1996–2012 (2002).

[4] M. R. Stinson, G. A. Daigle, and D. I. Havelock, ''The measurement of air-coupled surface waves over a porous surface,'' Proc. INTER-NOISE 93, pp. 1679–1682 (Leuven, 1993).

[5] M. R. Stinson, ''A note on the use of an approximate formula to predict sound fields above an impedance plane due to a point source,'' J. Acoust. Soc. Am. **98**, 1810–1812 (1995).

[6] S.-I. Thomasson, ''Reflection of waves from a point source by an impedance boundary,'' J. Acoust. Soc. Am. **59**, 780–785 (1976).

[7] R. Seznec and M. Bérengier, ''Effet du sol sur la propagation d'une onde sphérique,'' Acustica **53**, 269–280 (1983).

[8] K. B. Rasmussen, ''A note on the calculation of sound propagation over impedance jumps and screens,'' J. Sound Vib. **84**, 598–602 (1982).

[9] M. E. Delany and E. N. Bazley, ''Acoustical Properties of Fibrous Absorbent Materials,'' Appl. Acoust. **3**, 105–116 (1970).

Gas jets as ultrasonic waveguides

D. W. Choi, D. A. Hutchins,[a] and D. R. Billson
School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom

(Received 5 October 2002; revised 2 June 2003; accepted 4 June 2003)

The study of the propagation of ultrasound within a gas jet is extended to obtain waveguide effects, where the jet collimates the ultrasonic beam from a transducer within the flow. Two methods have been investigated to achieve this—cooling the gas within the air jet, and using a different gas whose acoustic velocity is lower than air. Cooling an air jet to a temperature less than that of the surrounding air produced a noticeable waveguide effect. In addition, studies have been carried out using other selected gases, such as carbon dioxide (CO_2) with a lower acoustic velocity than air, and helium (He) with a higher value. The former gas enhanced confinement of the ultrasonic beam, whereas the latter caused divergence. An ideal solution was found to be a CO_2/air mixture, which produced a well-collimated sound field along the axis, while limiting the excess attenuation of pure CO_2 gas jets. The effectiveness of the waveguide using mixed gas jets in producing images in air-coupled testing of solids is demonstrated. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1596169]

PACS numbers: 43.20.El, 43.20.Mv, 43.38.Bs [DEC]

I. INTRODUCTION

Water jets have been used in some applications as a transmission medium for ultrasound, and can be used in air to produce an effective waveguide.[1] This is caused by the very different acoustic properties that exist between water and air, including acoustic velocity and impedance. Several papers in this area have been published, and applications to measurements on solid surfaces, and for the through-transmission C-scanning of materials such as engineering composites, have been demonstrated.[1] Such transducers have advantages when rapid scanning of large areas is required, although the interaction of ultrasound with flowing water can be complicated, and the presence of water may be inconvenient. For these reasons, this paper describes a study into the waveguide effects of gas jets.

It is well known that a nonstationary medium can have a large effect on the diffraction field of an acoustic source. In the particular case of gas jets, there has been much work to study this phenomenon, especially in the context of noise reduction from aircraft engines.[2–6] Here, a ''cone of silence'' has been reported, which results from divergence of the emitted sound by refraction effects, so that the acoustic sound pressure levels along the axis reduces as energy is diverted away from the jet axis. This primary mechanism is thought to be the increased sound velocity within the jet due to gas flow away from the jet nozzle. Similar effects would be expected at much higher frequencies, and, in a previous paper,[7] it was observed that placing an ultrasonic transducer within an air jet at ambient temperature caused the resulting ultrasonic field to diverge, as expected from the increase in acoustic velocity within the gas jet with respect to the still air surrounding it.

However, for many applications where ultrasound is used, it would be beneficial to confine the ultrasonic beam within the gas jet, using the same refraction effects. The gas jet would then act as a waveguide. The confined beam would ideally have more improved directivity and lower attenuation, so that more effective imaging experiments could be performed with a better lateral resolution. In principle, this could be achieved by reducing the acoustic velocity within the gas jet to below ambient, so that acoustic energy would be confined within the gas jet, and an acoustic waveguide would be obtained.

In this paper, two methods of reducing the acoustic velocity within gas jets have been investigated: cooling the air jet emitted by a nozzle to below room temperature, and using a different gas in the jet whose acoustic velocity is lower than air. The first method has been reported in previous work. For example, focusing effects were noted in experiments at audible acoustic frequencies in very cold gas jets,[8] and previous work by the authors has demonstrated that effects can occur at ultrasonic frequencies.[7] In this paper, we are interested in both this approach and the use of different gases. Even though flow initially increases the acoustic velocity, the use of a suitable gas can be used to more than compensate for this. For example, the use of CO_2 gas whose acoustic velocity (259 ms^{-1}) is lower than that of air (331 ms^{-1} at the same temperature) would be expected to help contain the ultrasonic beam within the gas jet. Conversely, the acoustic velocity in a gas such as He is higher than that in air (970 ms^{-1} at 0 °C), and hence using He as a jet would be expected to increase divergence of the ultrasonic beam due to refraction.

The present paper studies the effects produced on the radiated field of an ultrasonic transducer, which is placed in gas jets of different gas composition. The results will be compared to those obtained by cooling. It will be demonstrated that effective collimation can be obtained. The result is potential for a gas-jet imaging system, where an increased resolution over conventional air-coupled imaging systems can be obtained.

[a]Author to whom correspondence should be addressed. Electronic mail: d.hutchins@warwick.ac.uk

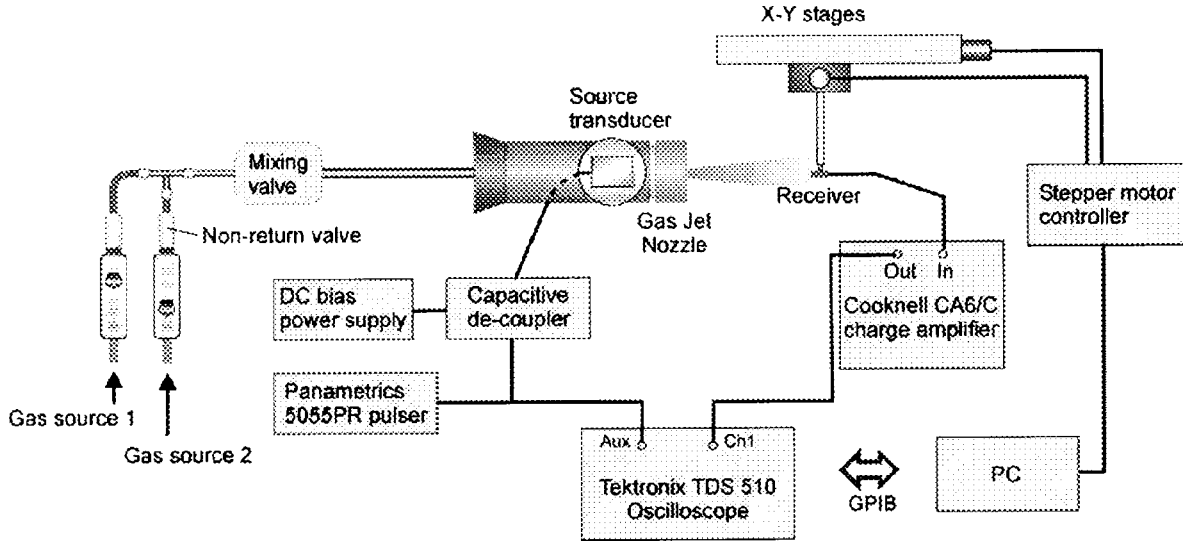


FIG. 1. Schematic diagram of a gas-jet transducer system, and the apparatus used to measure the radiated ultrasonic field.

II. EXPERIMENTAL ARRANGEMENT

The experiments were designed to measure the ultrasonic field emitted from a micromachined capacitive ultrasonic transducer, positioned axially within a gas jet. This gas jet was to be capable of being a single gas or a gas mixture. This was achieved by placing the transducer within a cylinder containing gas flow, and then fitting an appropriate nozzle in front of this to produce the gas jet. This arrangement was described in a previous paper.[7] The ultrasonic beam was emitted from this nozzle, symmetric axially with respect to the gas jet. This is shown schematically in Fig. 1, where the arrangement for mixing various gases at controlled flow rates was provided. Each source of gas (usually pressurized cylinders) had a regulator and a control valve. Gases passing through each flow measuring tube were prevented from flowing back using nonreturn valves fitted at the end of each flow tube. They then entered a mixing valve, so that the relative proportions of each gas could be adjusted.

The nozzle was of an ogive design,[9,10] chosen for the ability to produce extended regions of constant gas velocity in the emitted jet. The gas nozzle helps to accelerate the gas flow, and hence would help to smooth out any turbulence caused by the transducer within the pipe. To confirm that a gas jet with desirable properties resulted, a hot-wire anemometer was scanned through a jet formed from air at room temperature, with air being supplied at a rate of 5.0 l/min to the cylinder containing the transducer. As can be seen in Fig. 2, a well-behaved turbulent jet resulted, with a typical exit velocity at the nozzle of approximately 4.5 ms^{-1}. It will be seen that a relatively constant flow velocity region was produced close to the aperture of the nozzle, as expected from the ogive design. For a nozzle diameter of 5 mm, it was found that the exit air flow velocity could be varied conveniently from still air to 20 ms^{-1}. The gas jet itself had other properties that would be expected from fluid dynamics,[11–13] namely the creation of vortices at the boundaries to still air, and a divergence angle that increased with the exit velocity.

In certain measurements, the temperature of the gas jets was to be lowered to below room temperature. The additional apparatus to do this is shown in Fig. 3. It used a cooling unit situated between the gas sources and the pipe holding the ultrasonic transducer. The NESLAB immersion cooling unit consisted of a bath cooler (IBC-4A) and a controller (Croytrol). The bath had a cylindrical cooler with dimensions of 50 mm diameter and 110 mm length, which was submerged in fluid. This cooler was made of nickel-plated bronze and produced temperatures in the liquid down to −25 °C. The cooler was connected to the main body via a flexible insulated hose, which aided in positioning the probe within the cooling bath.

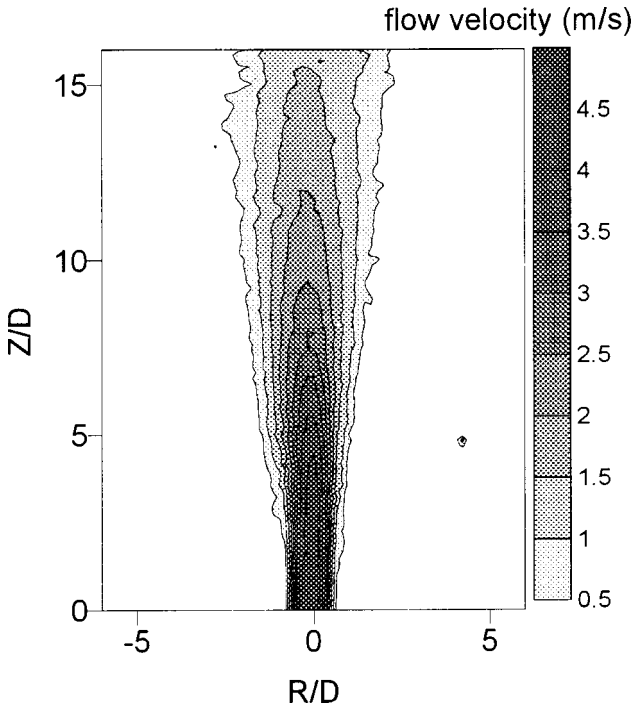


FIG. 2. Flow velocity field from a 5-mm ogive nozzle at an exit velocity of 4.5 ms^{-1}.

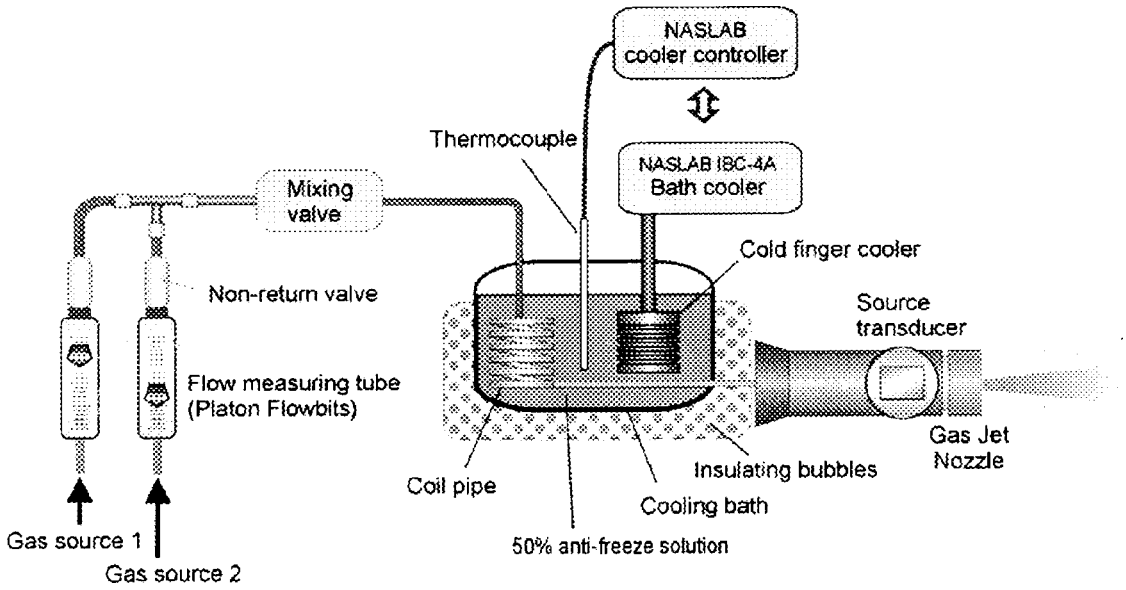


FIG. 3. Schematic diagram of the apparatus used to lower the temperature of gas jets.

The gas to be cooled passed through a coil of brass tubing, which was immersed in the cylindrical cooling bath. The bath contained a 50:50 mixture of water and ethylene glycol. Although the temperature of the liquid could be lowered to a preset value, the gas at the exit of the nozzle would not reach this temperature. This is because of ineffective heat transfer from the cooled brass and the gas, and heat transfer through the hose, the nozzle pipe, and so on. Therefore, to keep the system as cold as possible, it was found essential to insulate the apparatus carefully to minimize heat losses.

The source transducer was a capacitive device, fabricated using a thin metallized Mylar membrane against a micromachined silicon backplate. These transducers have been described extensively in the literature,[14,15] and their characteristics are well known. They give wide bandwidth signals at frequencies of up to 1 MHz. This was excited by voltage transients from a Panametrics 5055PR pulser, which was supplied to the transducer via a capacitive decoupler. This allowed dc bias voltages of typically 100 V to be applied to the device, to attract the membrane to the transducer, and to improve bandwidths. Spatial variations in the radiated ultrasonic fields were obtained by scanning cross-sections of the flow using a miniature capacitive receiver of 3-mm diameter. This was of a similar design to the source, and was connected to a Cooknell charge amplifier, which also provided the required 100 V dc bias signal for detection. Received waveforms were recorded on a Tektronix TDS510 oscilloscope and transferred to a PC for analysis and storage. The same PC controlled the stepper motor stages, as shown. Each field scan started at a distance of approximately 10 mm away from the source along its axis, over a scan area of 80 mm in the axial (Z) direction and 60 mm in radial (R) direction. Typically, scans were carried out at 1-mm spatial intervals, and the results were displayed in the form of either 3D surface plots or gray-scale images (sometimes the latter gave a better qualitative image for observing certain features). Distances were displayed in dimensionless distance units relative to the nozzle diameter D, namely as Z/D and R/D in the axial and radial directions, respectively.

As an illustration of the scanning procedure, Fig. 4 shows results from a scan of the ultrasonic field in still air, i.e., with no gas flow present. This acts as a reference field with which to compare the results with gas jets being present. Note that the field shows a steady decay in ampli-

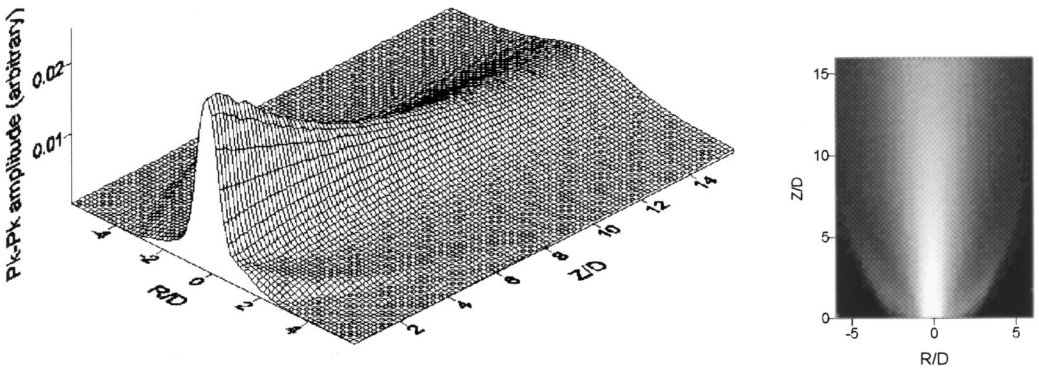


FIG. 4. Radiated ultrasonic field from the 5-mm ogive nozzle in still air.

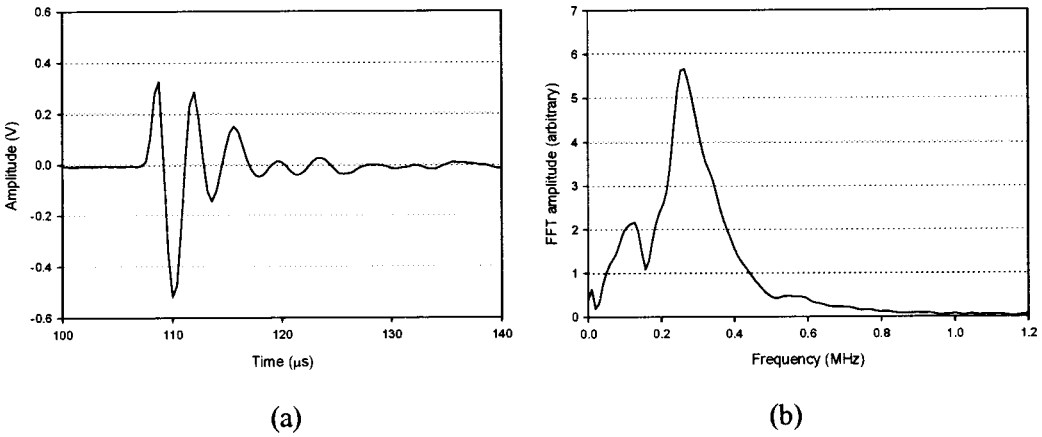


FIG. 5. (a) Typical ultrasonic waveform detected on-axis, and (b) its frequency spectrum via an FFT.

tude with distance from the nozzle, with little evidence of multiple reflection or other unwanted features. Figure 5 shows a typical waveform received on-axis, together with the frequency spectrum of the signal. The waveform exhibits a reasonably well-damped behavior, with a frequency spectrum showing peak amplitudes at ~300 kHz.

III. RADIATED SOUND FIELDS IN VARIOUS GAS JETS

The radiated ultrasonic fields in the presence of various gas jets were now investigated, and compared to the field that resulted in the absence of gas flow, shown earlier in Fig. 4. Initial experiments used various pure gases at room temperature. A constant input flow rate of 1 l/min was applied, which corresponded to an exit flow velocity of approximately 1.0 ms^{-1} for each measurement. Figure 6 shows a comparison of the radiated ultrasonic fields for air, He and CO_2. The results for an air jet [Fig. 6(a)] may be compared to that for still air, and it can be seen that the ultrasonic beam was relatively unaffected by air flow (note that this is not the case for higher exit gas velocities, as was seen in earlier work[7]). Conversely, the He jet caused the ultrasonic beam to diverge strongly due to the increased acoustic velocity within the gas jet, as would be expected [Fig. 6(b)]. The beam decayed rapidly in the axial direction, with most of the energy radiating at large angles to the axis (this is shown more obviously in the gray-scale image). This is a result of the large velocity difference between He and air, and demonstrates that refraction is the dominant feature causing energy loss in this case. This divergence of the acoustic beam when He is the jet gas also fits in with the concept of energy leakage or reradiation from faster to slower acoustic media, as for example in the case of leaky waves. This would not be observed for a slower gas jet acoustic velocity than air. Hence, as expected, the CO_2 gas jet [Fig. 6(c)] shows a reasonably collimated beam, but, as expected from measurements reported in the literature,[16] the attenuation at the ultrasonic frequencies used was very high in the pure gas. Hence, the signal decreased rapidly in the axial direction, but there was little evidence of divergence caused by refraction effects.

The measurements of Fig. 6 indicated that pure gases do not lead to very desirable properties in terms of waveguide effects—CO_2 exhibits too high an attenuation, and He causes dramatic refraction effects and loss of energy along the jet axis. It was thus decided to investigate mixtures of both CO_2 and He with air.

Various amounts of either CO_2 or He were now mixed with air, using the apparatus of Fig. 1, and the radiated ultrasonic field plotted. The resultant ultrasonic fields for a mixed gas jet of air and helium are shown in Fig. 7 for various He/air mixtures. This was achieved by keeping the velocity of gas at the exit nozzle constant, at a value of approximately 2.0 ms^{-1}, and varying the percentage of He gas in the mixture from 0% to 50% by volume. For pure air, Fig. 7(a), the increased flow rate has caused some divergence of the radiated ultrasonic beam, seen as the two farfield lobes at small angles to the Z axis. The addition of a small amount (17%) of He gas [Fig. 7(b)] caused this divergence to increase markedly, due to the higher acoustic velocity in the mixture. Further increases in the concentration of He caused more rapid axial decay, as more energy was radiated at larger angles [Figs. 7(c) and (d)]. This result is further evidence that an increased acoustic velocity does indeed cause refraction and hence energy loss at large angles. Gas jets of He are thus not suitable for the production of collimated ultrasonic fields, but the result may be of use in beam-steering applications.

The opposite effects would be expected in CO_2 gas jets, where the lower acoustic velocity has potential for use as a waveguide. However, as noted above, CO_2 gas has a higher attenuation coefficient than either air or He. In order to overcome the high sound attenuation of CO_2 gas, it was predicted that dilution of the gas with air, whose attenuation coefficient is relatively low, would be helpful. Figure 8 shows the results obtained for various volumetric ratios of CO_2 and air. With 12% of CO_2 gas by volume in the mixture, the result has improved slightly over the field measured in a pure air jet [Fig. 7(a)]; however, greater collimation effects are evident as the percentage of CO_2 gas in the mixture was increased. Values of 17% and 25% [Figs. 8(c) and (d)] produced very well-collimated fields, which seemed to be contained within the gas jet itself. While there is some evidence that the in-

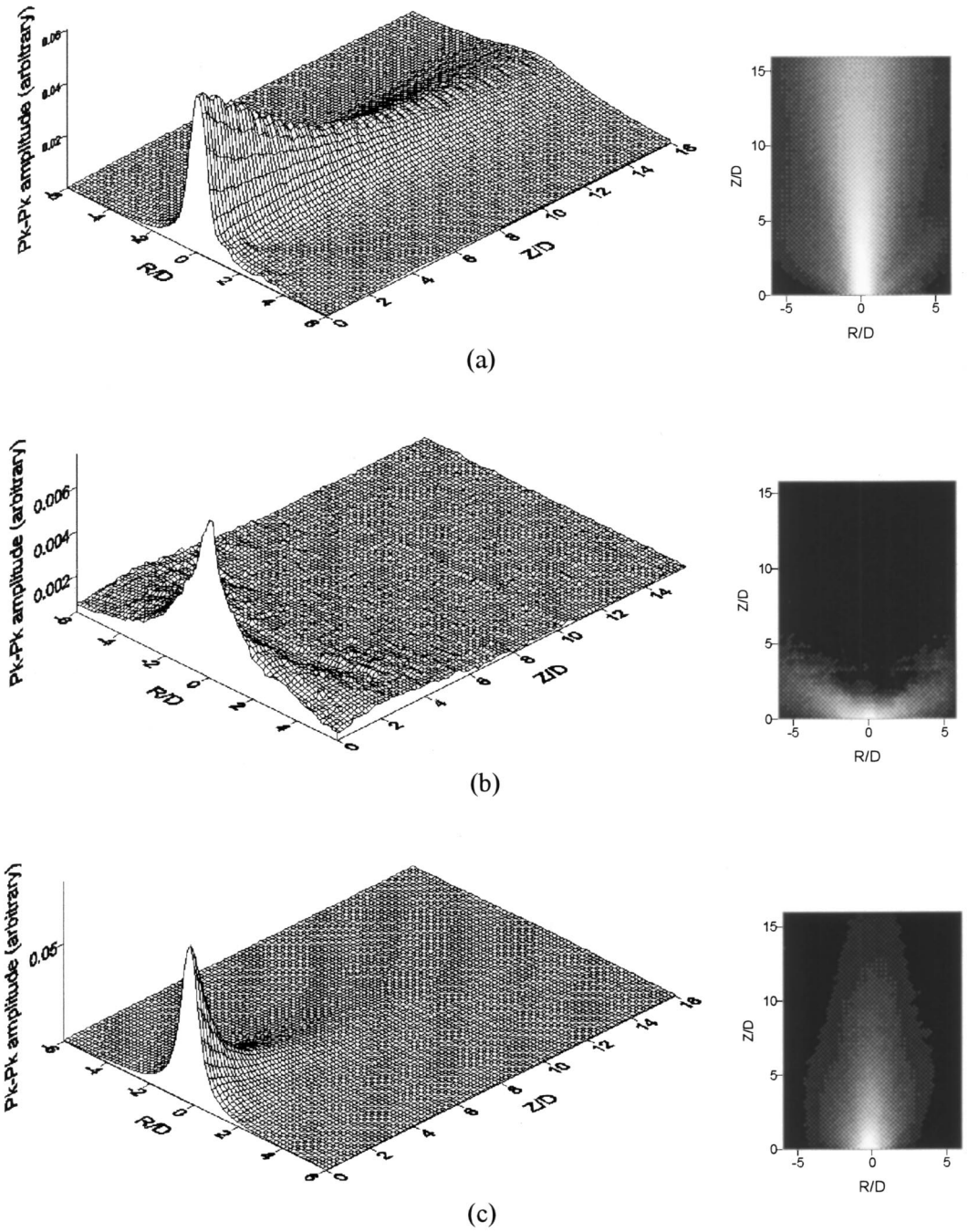


FIG. 6. Ultrasonic fields for an exit flow velocity of 1.0 ms^{-1}, using (a) air, (b) He, and (c) CO_2. (Images on the right-hand side are an alternative method of display, where a lighter shade represents an increase in amplitude.)

creased concentration of CO_2 had enhanced attenuation at larger axial distances, a maximum appeared in the axial peak-to-peak amplitude at larger CO_2 concentrations. It was also found that the position of this maximum in the radiated field moved further away from the nozzle as the percentage of CO_2 gas increased. This seems to be connected with the nearfield/farfield boundary of a piston source $z=a^2/\lambda$, where a is the radius of the source and λ is the wavelength. The distance in z of the nearfield/farfield boundary is inversely proportional to the acoustic velocity, and hence an increased percentage of CO_2 gas, leading to lower acoustic velocities, will cause the boundary to move away from the nozzle, as seen in Fig. 8(d). Note that this would not be visible in the field for pure CO_2 of Fig. 7(c), as the attenuation was too high.

It is interesting to note that the presence of CO_2 caused the divergence effects of gas flow to be suppressed, but there would still be an effect from gas flow trying to increase acoustic velocities in the Z direction, and hence cause divergence. It was thus thought interesting to study the effect of keeping the CO_2/gas mixture ratio constant at 17%, but to vary the total exit flow velocities between values of 2.5 and

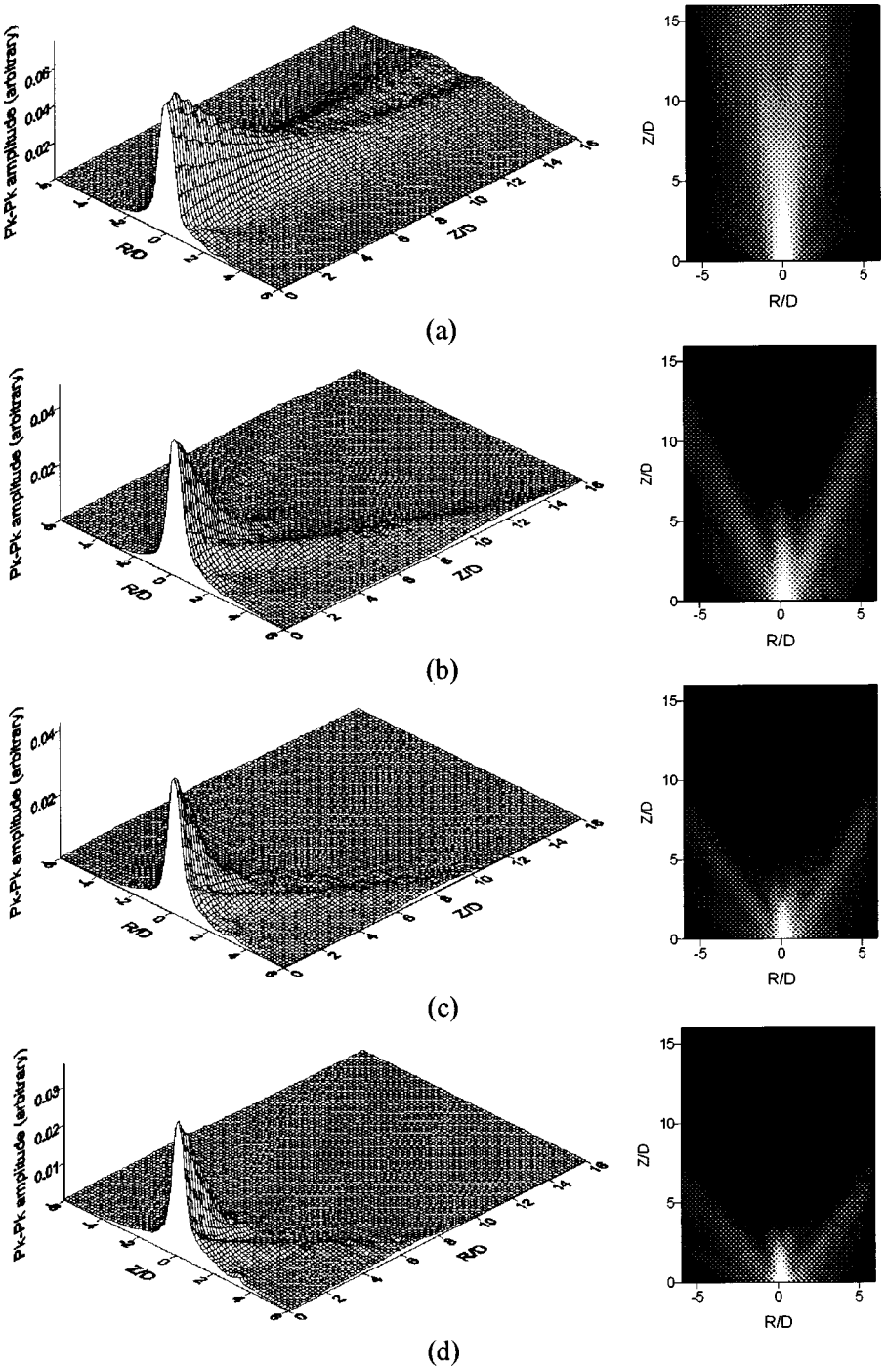


FIG. 7. Radiated ultrasonic fields at various He:air mixing ratios, obtained by changing the flow rate of He for a constant flow rate of air at 2.0 l/min. The concentrations of He were (a) 0%, (b) 17%, (c) 33%, and (d) 50%.

12 ms^{-1}. The results are shown in Fig. 9. As expected, decreasing the exit flow velocity led to greater collimation and less divergence due to refraction. It can thus be stated that the properties of these mixed gas jets are an interesting interaction between increases in the axial component of acoustic velocity due to gas flow and a decrease caused by adding CO_2 to the gas mixture. There is thus likely to be an optimum condition for a given measurement or imaging experiment. In the present experiments, this was thought to be an exit velocity from the 5-mm nozzle of 2.5 ms^{-1}, and a CO_2/gas mixture ratio constant at 17%.

In earlier work,[7] it was shown that lowering the temperature of a gas jet could lead to waveguide behavior. It was thus thought interesting to observe the effect of combining two methods, namely cooling mixed gas jets to below room temperature. A gas mixture of CO_2/air at 17% of CO_2 by volume [the same condition as Fig. 9(d)] was thus cooled to an exit temperature of $-0.5\,°C$, so that there was an approximate difference in temperature between the gas jet and the surrounding air of 23 °C. The resultant ultrasonic field in the cooled mixed gas jet is shown in Fig. 10. The field shows that the beam initially converged forming a narrow collimated zone, and then diverged as it propagated along the *Z* axis. This result is thought to be interesting as it indicates that a degree of focusing had occurred on-axis, as a result of too high a velocity difference between the gas jet and the surrounding air. Therefore, it was concluded that either a mixed gas jet, or cooling, both gave desirable effects, but

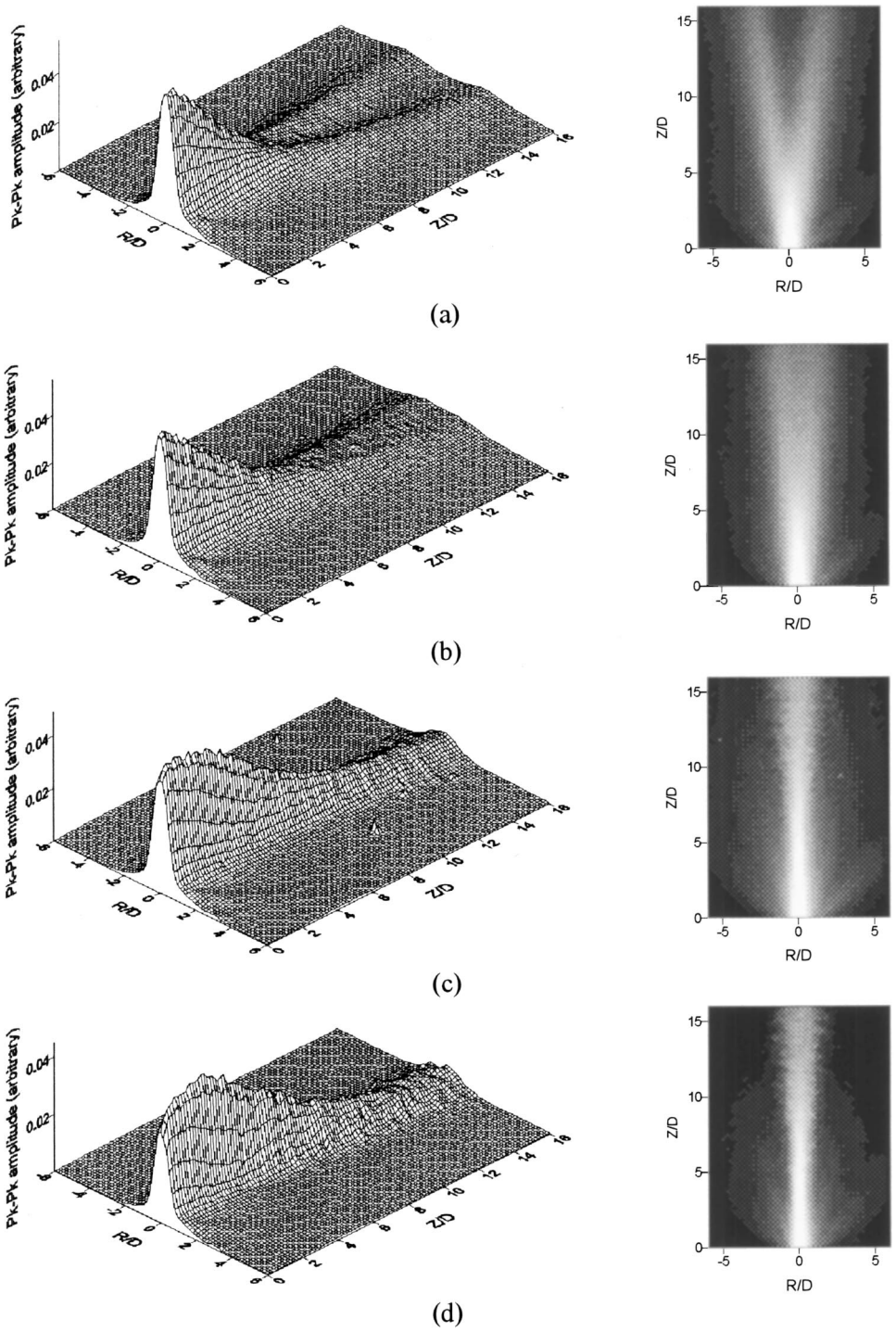


FIG. 8. Radiated ultrasonic fields at various CO_2:air mixing ratios, obtained by changing the flow rate of air for a constant flow rate of CO_2 gas at 1.0 l/min. The concentrations of CO_2 were (a) 9%, (b) 12%, (c) 17%, and (d) 25%.

that a jet which contained a cooled mixed jet was not necessary.

A final study investigated the use of different excitation waveforms with the capacitive transducer. In particular, it was thought interesting to study signals at lower frequencies, where diffraction effects would be expected to lead to a divergent beam; waveguide effects might be useful in helping to collimate such frequencies. Tone burst excitation at a lower frequency of 150 kHz was studied by replacing the Panametrics pulser in Fig. 1 with a Wavetek Model 191 waveform generator and an ENI RF power amplifier. The radiated sound field that resulted for this frequency in still air is shown in Fig. 11(a). The tone burst signal was then investigated for a mixed CO_2/air jet at 17% CO_2 by volume. The result, shown in Fig. 11(b), indicates that waveguide effects started to appear, but that at an increased CO_2 content of 23%, Fig. 11(c), greater collimation resulted. The better results at greater CO_2 concentrations were possible because of the lower attenuation in the gas mixtures at this frequency. The benefits of increased collimation could be obtained without the excessive signal losses that would result at higher frequencies. This is interesting, as it shows that this might be a method of producing highly directional

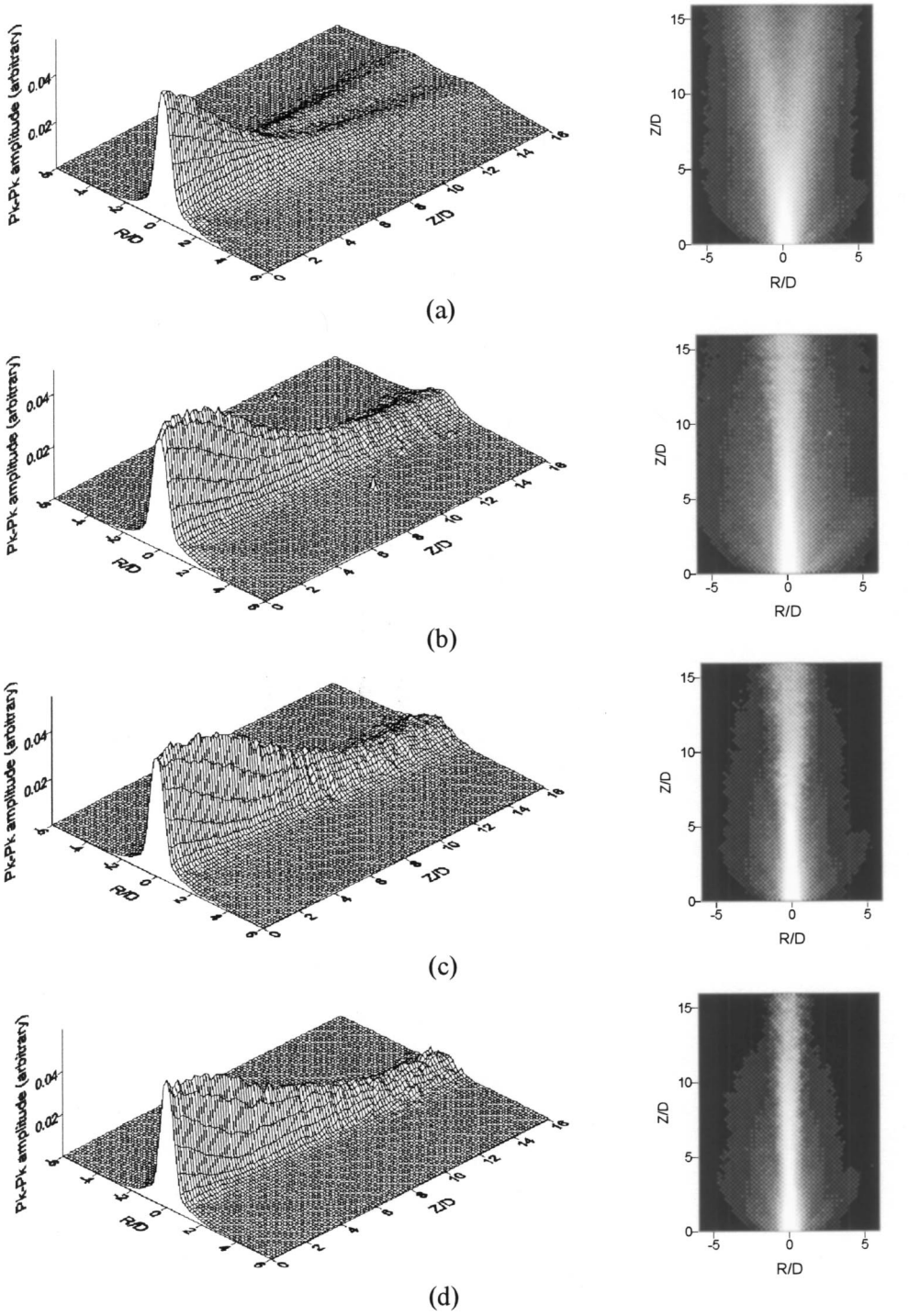


FIG. 9. Radiated ultrasonic fields of a gas jet with a 17% CO_2:air ratio, and exit gas velocities of (a) 12 ms^{-1}, (b) 6 ms^{-1}, (c) 3.5 ms^{-1}, and (d) 2.5 ms^{-1}.

beams of even lower frequencies. This is the subject of current work.

It is interesting to note the similarities of the above behavior, especially for broadband ultrasound in the captured gas jet column, with those of water jets. The filtering effects of a captured water column are well known, in that high frequencies tend to be lost. Here, similar effects are observed, but the cause is more likely to be the inherent attenuation within the gas itself, rather than any phenomena due to turbulence within the jet, or other waveguide effects. This is also more likely because of the lower frequencies used in the gas jets.

IV. SURFACE IMAGING USING GAS JET TRANSDUCERS

The gas jets have been shown to produce well-collimated ultrasonic beams. These could be expected to provide advantages in ultrasonic imaging, as divergence of the beam would be reduced if wave guide effects were present, leading to an improvement in lateral resolution. Hence, measurements were performed to test this hypothesis, using pulse-echo signals reflected from the surface of a solid sample. At an air/solid interface where a large acoustic impedance mismatch is present (true for many common solid

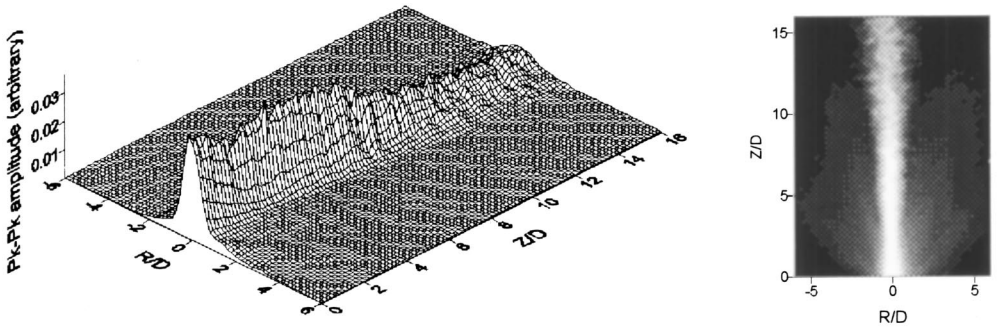
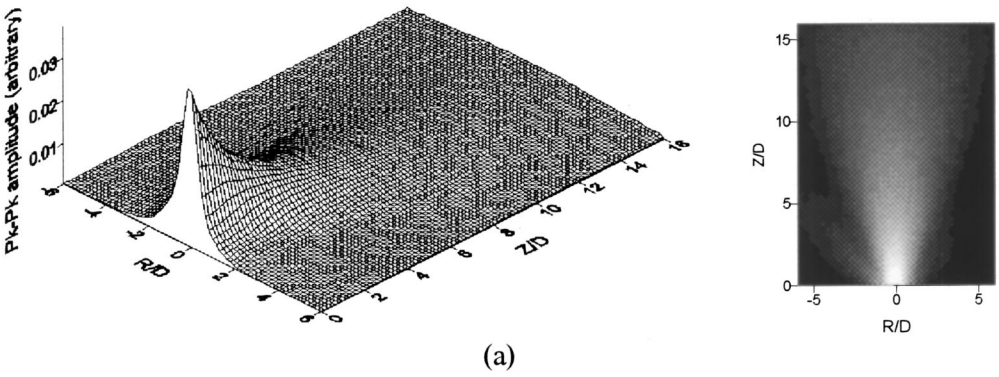
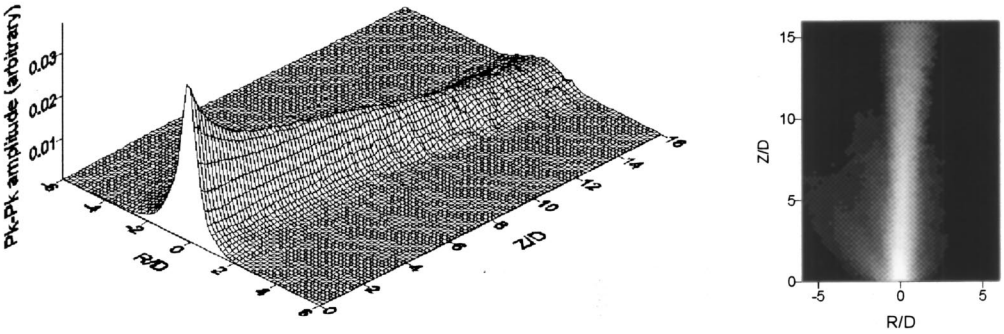


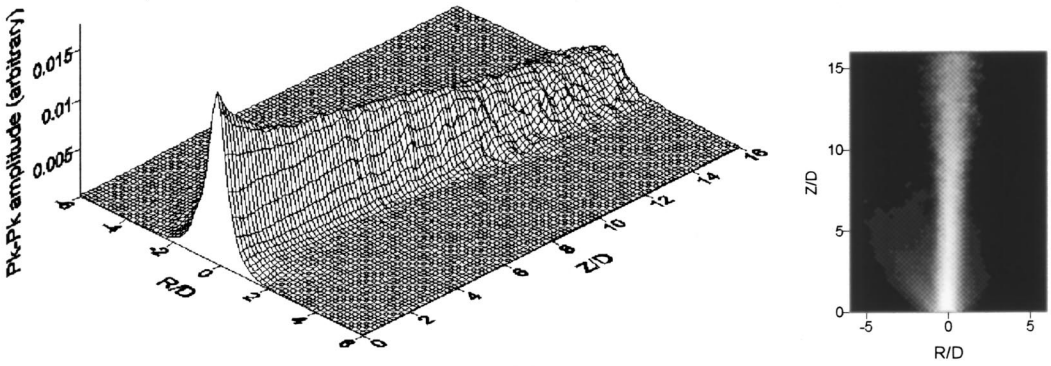
FIG. 10. Radiated ultrasonic field of 17% CO_2:air gas jet with an exit temperature of approximately -0.5 °C.



(a)



(b)



(c)

FIG. 11. Radiated ultrasonic fields for tone burst excitation at 150 kHz in (a) still air, (b) in a 17% CO_2:air jet, and (c) in a 23% CO_2:air jet.

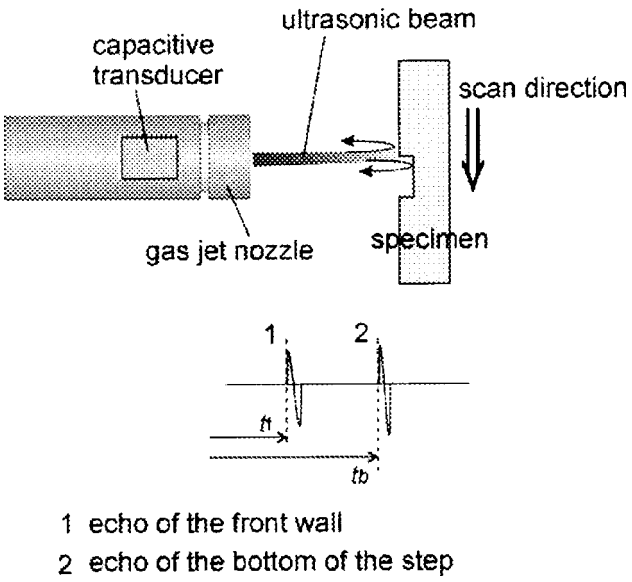


FIG. 12. Schematic diagram of a pulse-echo experiment.

materials in air), most of ultrasound signal will be reflected back to the source transducer, which can then be used to receive the signal, as shown in Fig. 12. A measurement of lateral resolution was performed by scanning this pulse-echo system across a step-change in height of an otherwise flat surface. A similar instrumentation system to the one shown earlier in Fig. 1 was used, with the receiving electronics con nected to the source, and the miniature scanned receiver removed. The scanning stages were now used to scan the ob ject under test past the transducer, acting in pulse-echo mode and driven with the Panametrics 5055PR pulser.

Data for a line scan across an 8-mm-deep step in a Plexiglas sample were recorded, and the time of flight extracted as a function of distance across the surface. This was performed at a distance of 30 mm between the nozzle and the surface at spatial intervals of 0.5 mm, and data were recorded for both still air and with a mixed CO_2/air jet at 17% of CO_2 by volume. Figure 13(a) shows the received normalized peak-to-peak amplitude steps as a function of *X*-position, and Fig. 13(b) shows the first-order spatial derivative, together with a fitted Gaussian curve. Note that the derivative is noisy due the nature of the data, hence a fitted curve is used. The full width of half-maximum amplitude (FWHM) of the Gaussian curve was taken as the nominal lateral resolution, as is conventional in imaging studies. The calculated FWHM of the step was approximately 5.4 mm for still air, and 2.9 mm when used. This demonstrates that, as expected, the use of an appropriate mixed gas jet has resulted in an improved lateral resolution. Note that for these measurements, a transient excitation signal was used, with the output waveform and spectrum as shown earlier in Fig. 5. The center frequency of these signals was thus 300 kHz.

To demonstrate the resultant improvement in resolution, a scan of a Plexiglas sample with a 20×20 mm^2, 8-mm-deep flat-bottomed hole was carried out, using a spatial measurement interval of 0.5 mm, over an area of 30×30 mm^2. Two sets of scans were performed, with the distance between the end of the nozzle and the front surface of the sample set at 15 or 30 mm. Figures 14 and 15 respectively show reflection images obtained in both still air and using a CO_2/air mixture at 17% of CO_2 by volume, at the two distances from the surface of the sample. Images were produced by taking the peak-to-peak amplitude of the first echo at each measurement point. Note that lighter areas represent higher amplitude and the darker areas lower amplitude. It can be seen that the image resolution is better in both cases when the gas jet was used, with less blurring of the image. This is a direct result of the improvements in lateral resolution that result from the use of gas jets.

V. CONCLUSIONS

It has been observed that the radiated field from an ultrasonic transducer can be modified by aligning it along the axis of a gas jet. This was achieved by placing the transducer within a cylinder containing gas flow, and fitting a subsequent ogive nozzle. This resulted in well-controlled jet characteristics. Previous work had observed that flow of air caused divergence of the emitted ultrasonic beam, due to increases in the axial component of acoustic velocity. This led to an acoustic velocity gradient in radial direction, and refraction then caused divergence of the beam. The present experiments with He, which has a much higher acoustic velocity than air, confirmed this hypothesis, as marked divergence was observed as the concentration of He within a mixed gas jet was increased. This may be of interest in applications where the radiated angle needs to be adjusted during a measurement. Effectively, a conical radiation field is produced, with very little propagation along the transducer axis. The angle of this cone was seen to vary with the concentration of He in the gas jet.

The use of a CO_2/air mixture was seen to lead to collimation of the beam, as the acoustic velocity within the gas jet could now be arranged to be lower that that in the surrounding air. As might be expected, increases in the concentration of CO_2 within the gas jet led to increased confinement of the ultrasonic energy within the gas jet, so that an effective waveguide was produced. This was seen to overcome the initial divergence caused by gas flow effects. It was found that at 17% of CO_2 in the gas mixture, an optimum mixture seemed to be produced as a compromise between better collimation and increased acoustic absorption within the CO_2 gas, which is much higher than in air alone. It was further demonstrated that cooling this mixture led to even greater effects, as this too caused a decrease in the local sound speed. The decreased divergence caused by these waveguide effects was shown to lead to increased lateral resolution, when the gas jet transducer assembly was used to image surface features in pulse-echo mode. This could possibly lead to applications in surface imaging, especially at lower acoustic frequencies, where collimation of acoustic beams becomes more difficult.

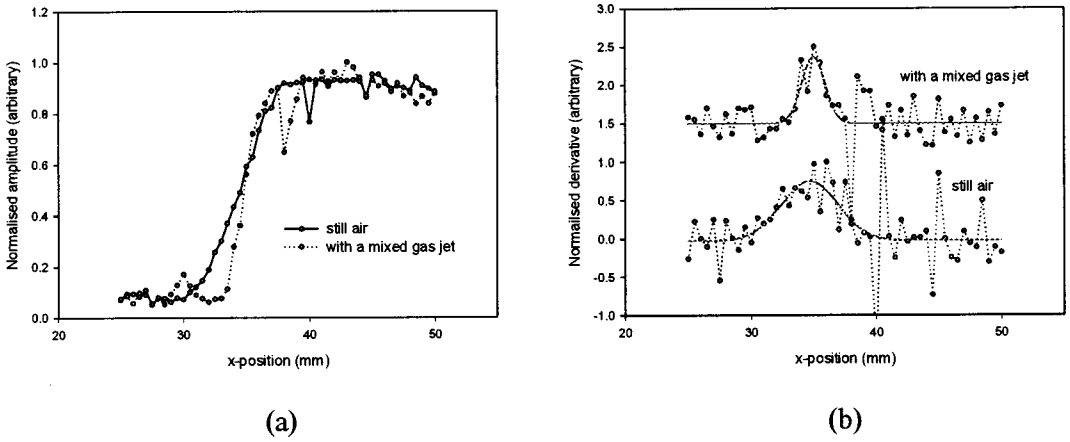


FIG. 13. Line scans across a step in the surface of a Plexiglas surface. (a) Received normalized peak-to-peak amplitude, and (b) first-order spatial derivative, together with fitted Gaussian functions. The jet nozzle was 30 mm from the upper sample surface.

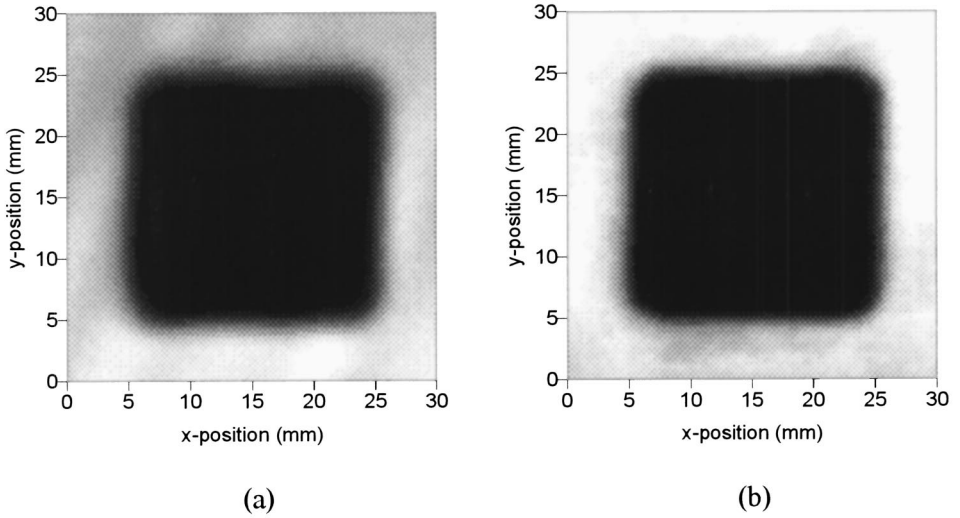


FIG. 14. Ultrasonic reflection images of a 20×20 mm^2 flat-bottomed hole in a Perspex plate (a) in a still air and (b) in a mixed air/CO_2 gas jet, at a distance of 15 mm.

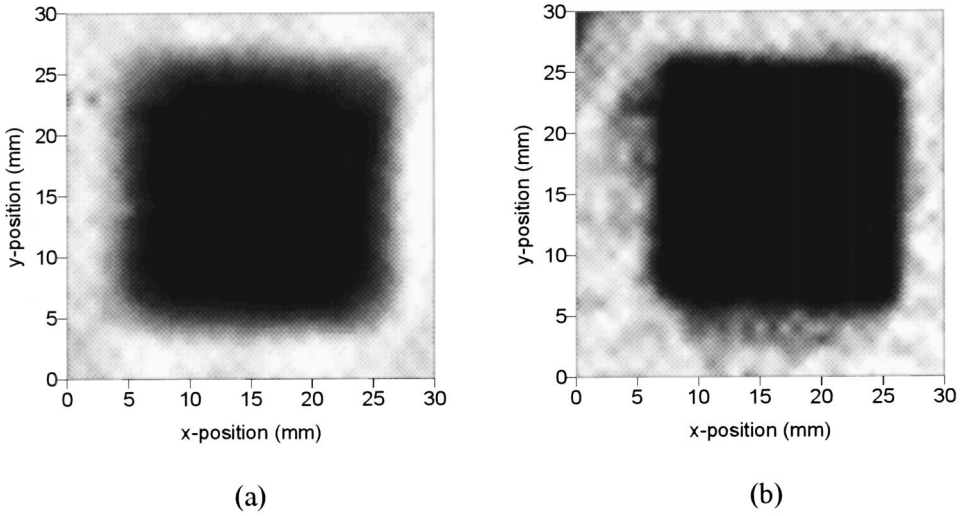


FIG. 15. As in Fig. 14, but at a distance of 30 mm.

ACKNOWLEDGMENTS

This work was supported via a grant from the Engineering and Physical Sciences Research Council.

[1] G. V. Blessing and D. G. Eitzen, ''Ultrasonic sensor for measuring surface roughness,'' Proc. SPIE **1009**, 281–289 (1988).
[2] G. M. Lilley, ''The radiated noise from isotropic turbulence with applications to the theory of jet noise,'' J. Sound Vib. **190**, 463–476 (1996).
[3] C. Bailly, C. Candel, and P. Lafon, ''Prediction of supersonic jet noise from a statistical acoustic model and a compressible turbulence closure,'' J. Sound Vib. **194**, 219–242 (1996).
[4] L. M. B. C. Campos, ''On Waves in Gases. 1. Acoustics of Jets, Turbulence and Ducts,'' Rev. Mod. Phys. **58**, 117–128 (1986).
[5] R. Vedantham and J. Hunter, ''Sound Wave Propagation through Incompressible Flows,'' Wave Motion **26**, 318–328 (1997).
[6] J. Atvars, L. K. Schubert, and H. S. Ribner, ''Refraction of sound from a point source placed in an air jet,'' J. Acoust. Soc. Am. **37**, 168–170 (1965).
[7] D. A. Hutchins, C. S. McIntyre, D. W. Choi, D. R. Billson, and T. J. Robertson, ''The propagation of ultrasound within a gas jet,'' J. Acoust. Soc. Am. **110**, 2964–2973 (2001).
[8] E. Grande, ''Refraction of injected sound by a very cold nitrogen jet,'' J. Acoust. Soc. Am. **38**, 1063–1064 (1965).
[9] R. J. Lamont, ''Pressures around an inclined ogive cylinder with laminar, transitional or turbulent separation,'' AIAA J. **20**, 1492–1499 (1982).
[10] C. K. W. Tam, ''Influence of Nozzle Geometry on the Noise of High-Speed Jets,'' AIAA J. **36**, 1396–1400 (1998).
[11] *Visualized Flow*, compiled by The Japan Society of Mechanical Engineers (Pergamon, New York, 1988).
[12] T. E. Faber, *Fluid Dynamics for Physicists* (Cambridge U.P., Cambridge, 1995).
[13] H. Schlichting and K. Gersten, *Boundary Layer Theory*, 8th ed. (McGraw–Hill, New York, 2000).
[14] D. W. Schindel, D. A. Hutchins, L. Zou, and M. Sayer, ''The Design and Characterization of Micromachined Air-Coupled Capacitance Transducers,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **UFFC-42**, 42–50 (1995).
[15] D. A. Hutchins, D. W. Schindel, A. G. Bashford, and W. M. D. Wright, ''Advances in Ultrasonic Electrostatic Transduction,'' Ultrasonics **36**, 1–6 (1998).
[16] D. R. Lide, Ed., *CRC Handbook of Chemistry and Physics*, 80th ed. (CRC, Boca Raton, FL, 1998).

Forward scattering of pulses from a rough sea surface by Fourier synthesis of parabolic equation solutions

David A. Miles,[a] Robin N. Hewitt, Marcus K. Donnelly, and Timothy Clarke
QinetiQ, Winfrith Technology Centre, Dorset DT2 8XJ, United Kingdom

(Received 11 July 2000; revised 22 January 2003; accepted 28 May 2003)

A variable depth step implementation of the range-dependent acoustic model (RAM) is applied to the modeling of forward scattering from a rough sea surface. The sea surface is treated within RAM simply as an internal interface between a water layer and an air upper halfspace. A comparison with a numerically exact integral equation is undertaken for the scattering of single frequencies from Pierson–Moskowitz sea surfaces. The method is extended to model the variability of linear frequency modulated pulses from a series of frozen sea surfaces in a shallow water waveguide. The subsequent effect of rough boundary scattering on the replica correlation process is investigated. [DOI: 10.1121/1.1595106]

PACS numbers: 43.20.Fn, 43.30.Bp, 43.30.Hw [DLB]

I. INTRODUCTION

The modeling of acoustic scattering from a rough sea surface boundary is a challenging problem. The application of approximate analytical methods to the problem of rough boundary scattering is possible, but these are based on either perturbation methods in which the acoustic wavelength is large compared to the surface features, or ray methods in which the reverse is true.[1] The region of particular interest is that in which the acoustic wavelength and sea surface features are of comparable size where neither of the above methods are sufficiently accurate. Holford[1,2] contributed to laying the groundwork for direct numerical solutions in this middle ground, though he specializes in periodic solutions. The presence of bubble clouds produced by the motion of the sea surface can further complicate the problem. For the frequencies considered in this paper bubble clouds have been shown to be important in the propagation of sound.[3] However, in this initial investigation, their presence is neglected and only the effects of a rough sea surface boundary are modeled.

Since the speed of the acoustic waves through the medium is much larger than the time derivative of the sea surface elevation it is a reasonable approximation to neglect the motion of the interface when considering its interaction with the acoustic wave.[4] This implies that the moving sea surface can be treated as a series of instantaneously fixed surfaces. However, predicting the scattering in any particular direction is still a complex problem. To reduce the computational complexity the backscattering and scattering out of the plane of propagation are neglected. Under these conditions the solution of acoustic scattering from a rough sea surface boundary is within the confines of the cylindrically symmetric three-dimensional (3-D) parabolic equation (PE) method.[5]

Work on the application of the PE method to rough surface scattering is available in the open literature. Dozier[6] employs conformal mapping to transform a section of the rough sea surface to a flat surface. This idea has been implemented in the finite element based program FEPE.[7] More recently, Rosenberg[8] has modified the range-dependent acoustic model[9–11] (RAM) by directly enforcing a pressure release boundary condition on a one-dimensional (1-D) rough surface. The approach taken by the present authors also utilizes the RAM model. However, the representation of the rough sea surface is implemented as an internal range-dependent interface between a water layer and an air upper halfspace. Thus, the rough sea surface is approximated as a cylindrically symmetric surface consisting of a sequence of stair-steps. RAM is particularly suited to this implementation as the interface conditions of continuity of pressure and normal particle displacement velocity are not explicitly applied: the Galerkin depth-discretization scheme[12] treats the internal interface as simply a rapid linear transition between the two layers. Hence the complexity involved in explicitly implementing the pressure-release condition on a range-dependent boundary can be avoided. The authors believe that this is a new approach to the representation of a rough sea surface in acoustic propagation problems.

An important enhancement to RAM for this research has been the inclusion of a variable computational depth step scheme. A fine depth step grid, used within a few acoustic wavelengths from an interface, is chosen to be an order of magnitude smaller than a coarse depth step grid that is used elsewhere. The variable depth step scheme is derived independently of that of Rosenberg,[8] but it follows a very similar line of reasoning and identical equations are obtained. A comparison of the efficiency of the variable and fixed depth step implementations of RAM is not undertaken. However, it is anticipated that the speed of the variable depth step implementation (using a fine and a coarse grid) will be an order of magnitude faster than the original fixed depth step version of RAM with a fine depth step grid. From this point onwards the authors adopt the convention that RAM(v) refers to the variable depth step implementation of the original RAM code.

The simulation of pulse propagation is achieved by using the Fourier synthesis technique. This entails combining single frequency solutions of the RAM(v) model with the

[a]Electronic mail: dmiles@QinetiQ.com

fast Fourier transform (FFT) to evaluate the spectrum and pulse at a receiver some distance from the source. This approach is implemented by the authors in a model known as synthetic pulse reception (SPUR).

A brief description of the Fourier synthesis technique is given in Sec. II. Details of the rough sea surface representation and various test case comparisons of the RAM(v) implementation with solutions of the numerically precise Helmholtz integral equation are given in Secs. III and IV, respectively. Finally, in Sec. V, SPUR and RAM(v) are utilized in the investigation of pulse variability due to a changing rough sea surface boundary in a shallow water environment. The subsequent effect of the rough sea surface on the replica correlation process is analyzed.

II. THE FOURIER SYNTHESIS METHOD

The effect of the ocean environment upon an acoustic signal can be described as a linear filter, the coefficients of which are obtained from spatially dependent Green's functions. Equations (1) and (2) show the time domain and frequency domain representations, respectively,

$$s'(\mathbf{r},t)=\int_{-\infty}^{\infty} g(\mathbf{r};\mathbf{r}_s,\tau)s(t-\tau)\,d\tau, \tag{1}$$

$$S'(\mathbf{r},\omega)=G(\mathbf{r};\mathbf{r}_s,\omega)S(\omega), \tag{2}$$

where s (or S) is the source signal in time (or frequency), s' (or S') is the received signal in time (or frequency), g (or G) is the Green's function in time (or frequency), t and τ are time, ω is the angular frequency, and $\mathbf{r}$ is a position vector (subscript s indicates source).

From Eq. (2), the inverse Fourier transform is used to obtain the received signal. It is evident from the time domain filter approach of Eq. (1) that the coefficients of the filter may vary with time due to temporal changes in the ocean. It is well established in the literature that there is a formal duality between the time and frequency domain approaches;[13] either approach can be taken and still account for time variations. In particular, temporal changes in the ocean can be modeled in the frequency domain by obtaining a set of transfer functions for a series of temporal ''snapshots'' of a time-varying ocean and using inverse FFTs to give a series of sequential filters. A time-dependent filter can then be derived by simple linear interpolation between these snapshot filters. In this paper the Doppler effects arising from the changes in the shape of the sea surface during the pulse travel time are neglected. However, the shape changes between pulse transmissions are modeled. In this manner a series of snap-shots of the sea surface are generated for each of the transmitted pulses and for each pulse the sea surface is considered frozen.

A common problem in Fourier synthesis methods in lossy media is ensuring causality of the resulting propagating pulse. This is due to the fact that in order to comply with causality an attenuating medium must be dispersive. The dispersion is related to the frequency-dependent attenuation through the Kramers–Kronig relations.[14] Kulkarni *et al.*[15] have made numerical comparisons between pulses propagating in attenuating ocean environments, with and without sediment dispersion. At the frequencies they consider (~100 Hz), the changes in the sound speed are significant and discrepancies in the modeling are apparent. For the pulse propagation modeling undertaken in the present paper, a value for the change in the sound speed over the pulse bandwidth (using the dispersion relation contained in Ref. 15) is found to be less than 2.5 ms^{-1}. The resulting errors in the magnitudes of the pressures at frequencies corresponding to the upper and lower ends of the pulse bandwidth are of the order of 1%, with corresponding errors in the phase of 1°. Therefore, the effects of dispersion are neglected and the sound speed in the halfspace is taken to be constant over the pulse bandwidth.

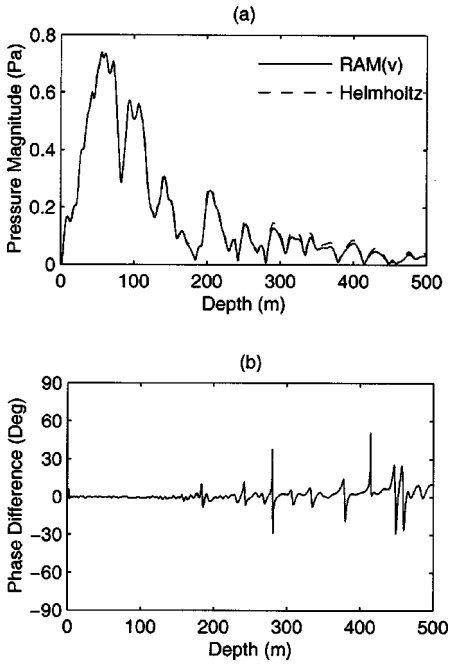


FIG. 1. Comparison between the Helmholtz and RAM(v) solutions for a Pierson–Moskowitz surface ($\nu=15\ \mathrm{ms}^{-1}$) at a frequency of 400 Hz; (a) pressure magnitude and (b) phase difference.

III. ROUGH SEA SURFACE REPRESENTATION

The existing PE method cannot model the full scattering solution of an acoustic wave from a two-dimensional (2-D) asymmetrically rough sea surface. A typical approximation often adopted in ocean acoustic modeling involves replacing the azimuthally dependent 2-D spectral density by an isotropic 1-D spectral density,[16] with the same root mean square (rms) height. This method is utilized throughout the present calculations. The 1-D sea surface thus generated is distinct from that obtained by taking a 1-D slice through the original 2-D sea surface but it demonstrates the important scattering effects due to the rough sea surface. A brief discussion of the derivation of a 1-D spectrum is given but the reader is referred to Thorsos[16,17] for a more rigorous description.

A realistic rough 2-D sea surface is described by a spectral density $S(\mathbf{k})$, where $\mathbf{k}$ is the sea surface wavenumber. The average value of the square of the sea surface height, h, is given by

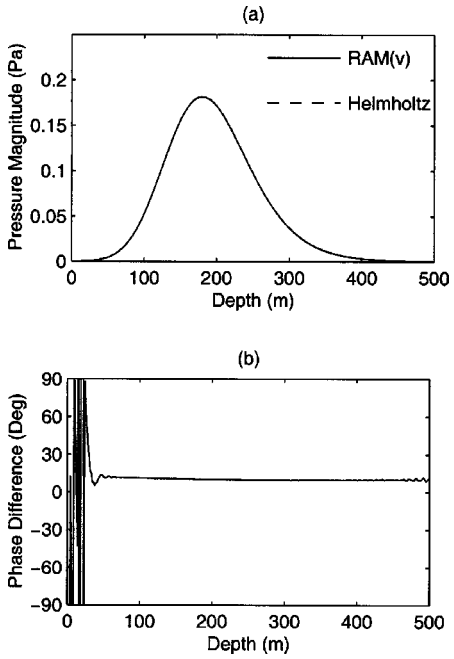


FIG. 2. Comparison between the Helmholtz and RAM(v) solutions for a flat surface at a frequency of 1500 Hz; (a) pressure magnitude and (b) phase difference.

$$\langle h^2\rangle = \int S(\mathbf{k})\, d\mathbf{k} = \int_0^{2\pi}\int_0^{\infty} S(k)\Theta(k,\alpha)k\, dk\, d\alpha, \qquad (3)$$

where k is the wave number magnitude, α is the angle between the average direction of the wind and the direction of propagation of a surface wave with wave vector $\mathbf{k}$, and $\Theta(k,\alpha)$ is the normalized angular distribution of the sea

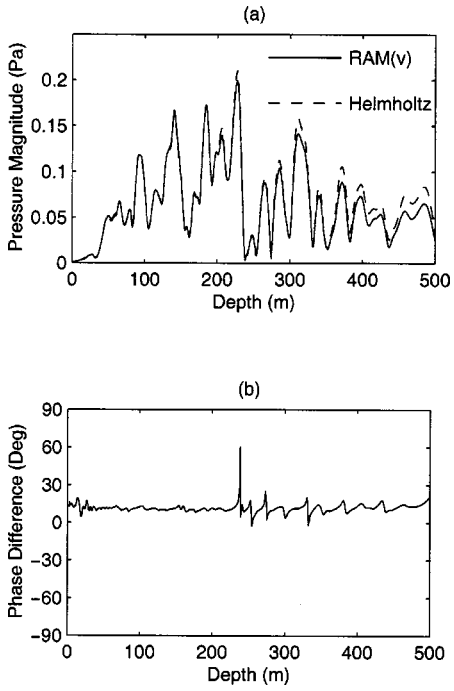


FIG. 3. Comparison between the Helmholtz and RAM(v) solutions for a Pierson–Moskowitz surface ($\nu = 15\ \mathrm{ms}^{-1}$) at a frequency of 1500 Hz; (a) pressure magnitude and (b) phase difference.

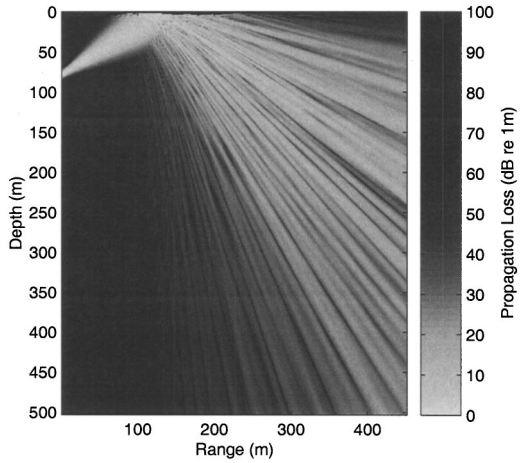


FIG. 4. Propagation loss versus depth and range for a Pierson–Moskowitz surface ($\nu = 15\ \mathrm{ms}^{-1}$) at a frequency of 1500 Hz generated with RAM(v).

wave energy. If an isotropic angular distribution is assumed, then Eq. (3) can be written as

$$\langle h^2\rangle = \int_0^{\infty} kS(k)\, dk = \int_0^{\infty} \overline{S}(k)\, dk, \qquad (4)$$

where $\overline{S}(k)$ is the equivalent 1-D isotropic sea surface spectrum.

Since a simple shallow water 2-D spectral density is unknown to the authors, the Pierson–Moskowitz[18] deep water spectral density is adopted. It is an empirical expression for wind-blown waves acting over an infinite fetch and is defined by

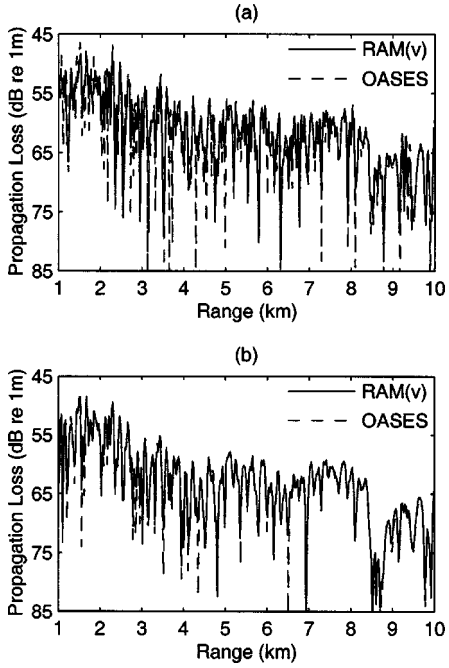


FIG. 5. Comparison between OASES and coherently averaged RAM(v) solutions at a frequency of 400 Hz at a receiver depth of 80 m; (a) flat sea surface and (b) Gaussian rough sea surface spectra.

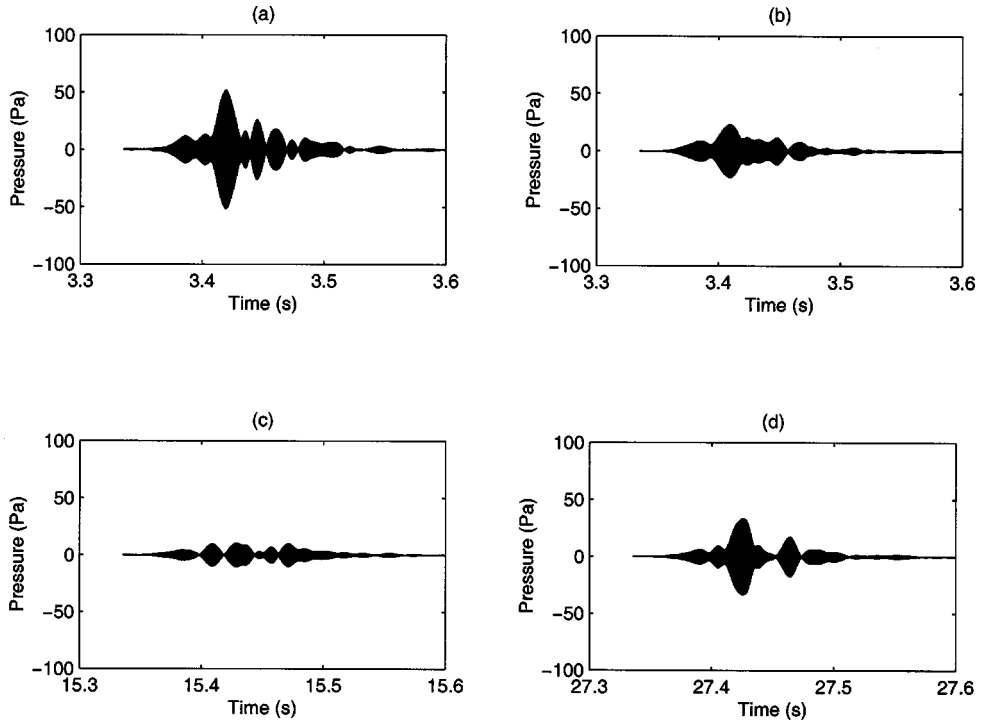


FIG. 6. Received pulses in a summer environment at a receiver depth of 10 m; (a) flat surface (b) rough sea surface at $t=0$ s, (c) rough sea surface at $t=12$ s, and (d) rough sea surface at $t=24$ s.

$$S(k)=8.1\times10^{-3}\times\frac{1}{2k^4}\exp\left(-\frac{0.74g^2}{\nu^4k^2}\right) \tag{5}$$

where ν is the wind speed in ms^{-1} measured at 19.5 m above the mean sea level and g is the acceleration due to gravity ($9.81\ \mathrm{ms}^{-2}$).

A brief explanation of how to obtain sea surface realisations is presented here in terms of the continuous variable k. This discussion does not include all of the details (which includes treating k as a discrete variable) and the reader is referred to Thorsos[17] for a more precise account.

Different realizations of the time-dependent sea surface height, $h(x,t)$, are obtained by applying normal probability distributions to the amplitude spectrum. Thus,

$$\sqrt{\overline{S(k)}}\Rightarrow\sqrt{\overline{S(k)}}\times\left\{\frac{N_1(k)+iN_2(k)}{\sqrt{2}}\right\} \tag{6}$$

[where $N_1(k)$ and $N_2(k)$ are random variates selected from a standard normal distribution] and subsequently taking the Fourier transform of $\sqrt{\overline{S(|k|)}}/2$ as defined in Eq. (7),

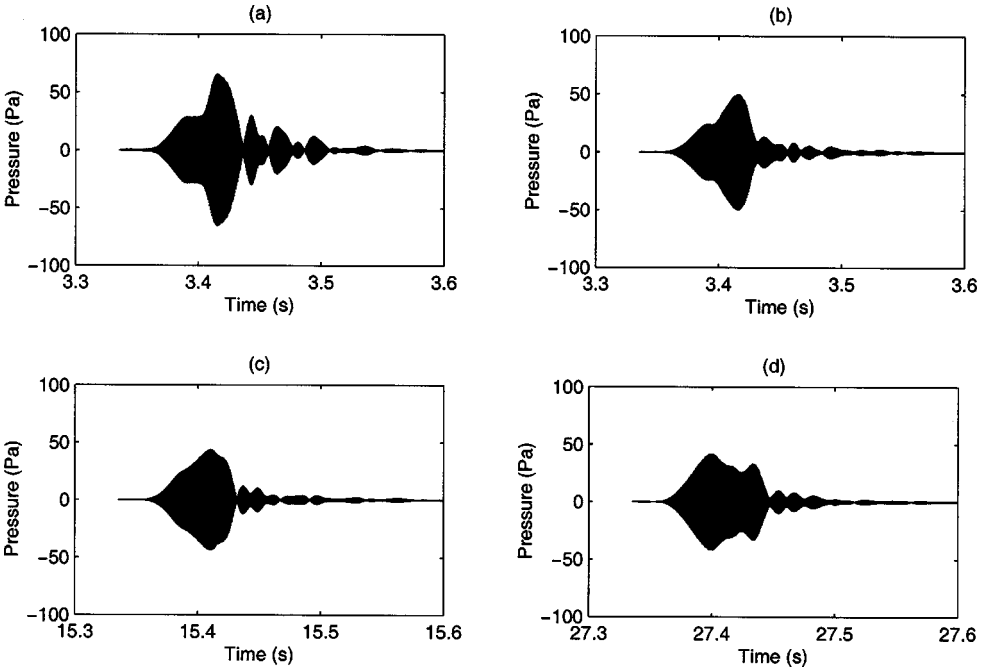


FIG. 7. Received pulses in a summer environment at a receiver depth of 35 m; (a) flat surface, (b) rough sea surface at $t=0$ s, (c) rough sea surface at $t=12$ s, and (d) rough sea surface at $t=24$ s.

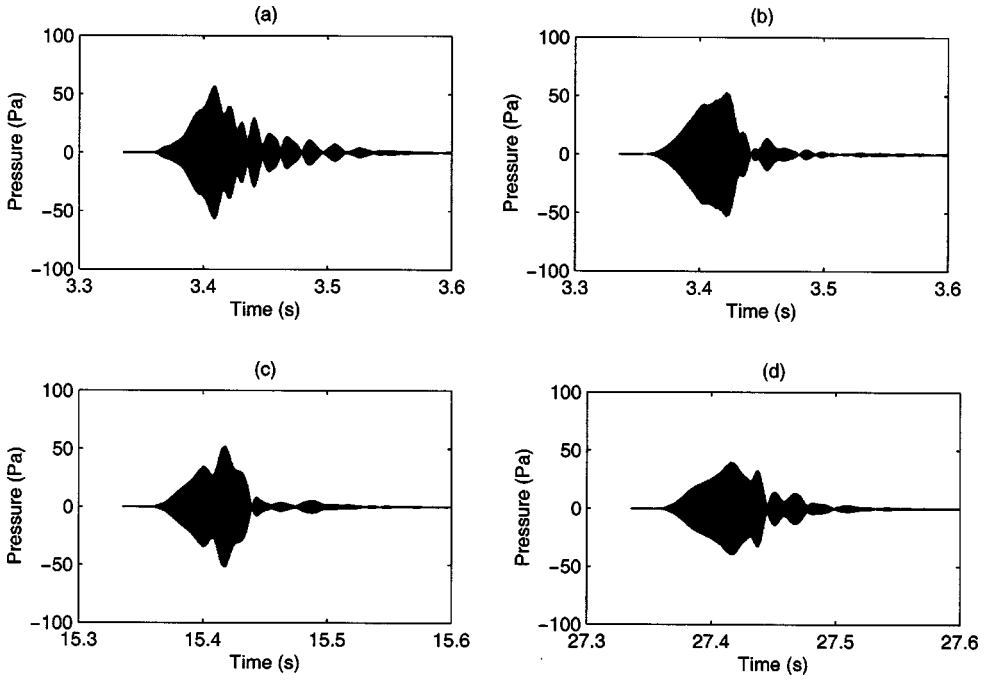


FIG. 8. Received pulses in a summer environment at a receiver depth of 75 m; (a) flat surface, (b) rough sea surface at $t=0$ s, (c) rough sea surface at $t=12$ s, and (d) rough sea surface at $t=24$ s.

$$h(x,t)=\int_{-\infty}^{\infty}\sqrt{\frac{\overline{S(|k|)}}{2}}\frac{(N_1(k)+iN_2(k))}{\sqrt{2}}e^{i(kx-\Omega t)}\,dk, \tag{7}$$

where Ω is the angular frequency of the sea surface waves. In reality the infinite interval is replaced by a finite interval when Eq. (7) is made discrete and $h(x,t)$ becomes periodic in x.

For consistency, the Pierson–Moskowitz spectrum is used together with the deep water dispersion relation as defined by Eq. (8):

$$\Omega^2=gk. \tag{8}$$

IV. REFERENCE SOLUTIONS FOR THE SEA SURFACE SCATTERING PROBLEM

Although RAM conserves energy by solving for a modified pressure variable, the interfaces dealt with are normally the seafloor and interfaces between different subseafloor layers. For such cases RAM has shown excellent agreement with benchmark solutions, such as those from two-way coupled mode solutions. However, the air–water interface has a much larger impedance mismatch. In order to verify the RAM(v) implementation and the treatment of the air–water boundary as an internal interface, benchmark solutions are produced using the numerically precise Helmholtz inte-

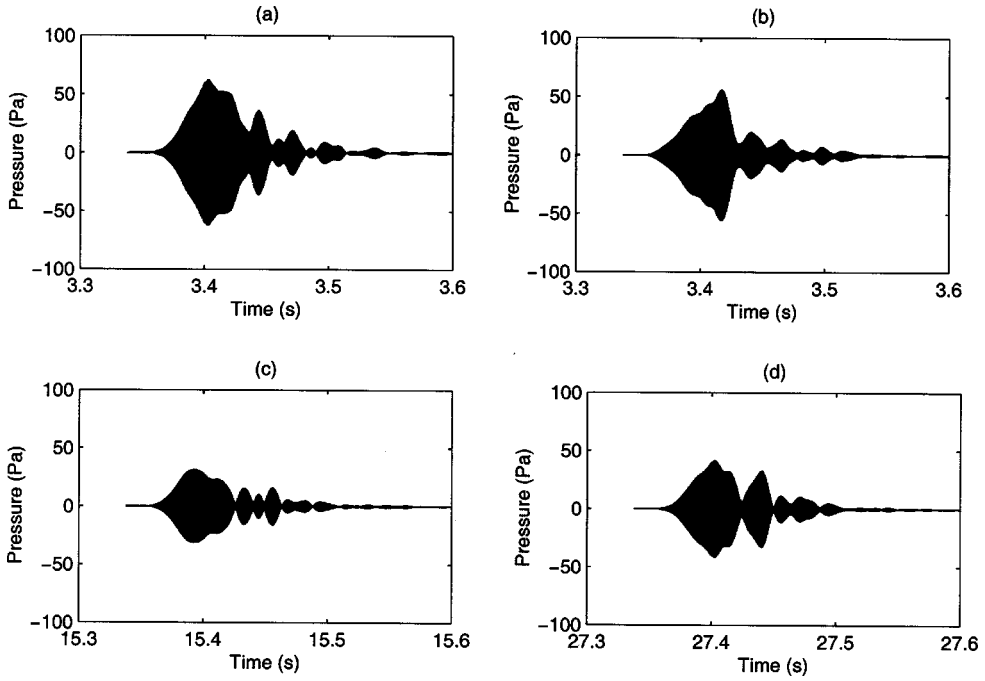


FIG. 9. Received pulses in a winter environment at a receiver depth of 10 m; (a) flat surface, (b) rough sea surface at $t=0$ s, (c) rough sea surface at $t=12$ s, and (d) rough sea surface at $t=24$ s.

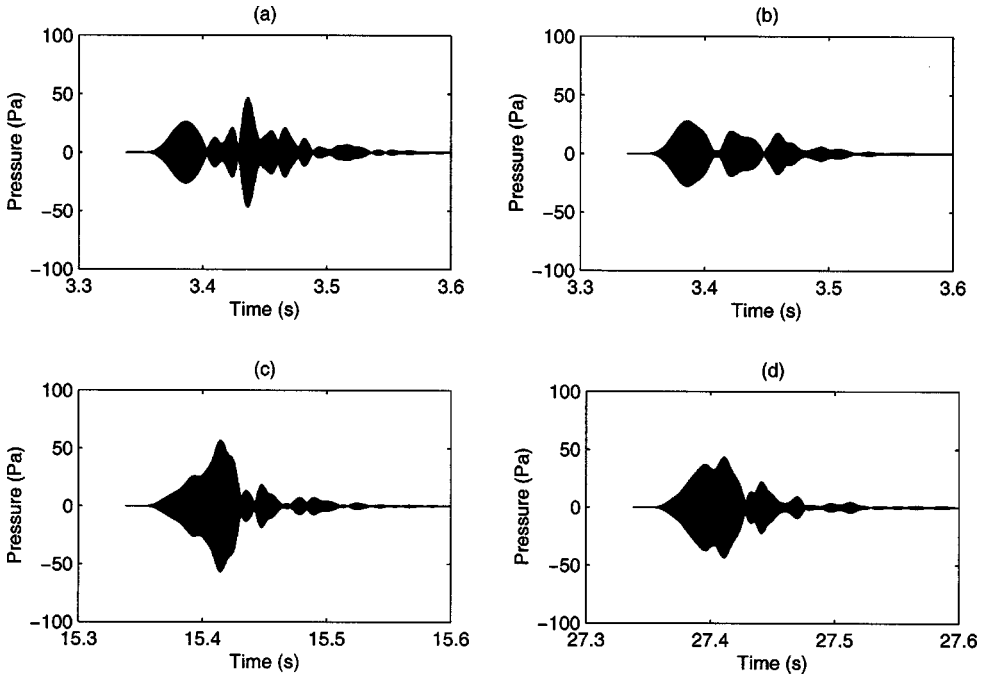


FIG. 10. Received pulses in a winter environment at a receiver depth of 35 m; (a) flat surface, (b) rough sea surface at $t=0$ s, (c) rough sea surface at $t=12$ s, and (d) rough sea surface at $t=24$ s.

gral method. The RAM(v) convergence parameters and CPU times relevant to the benchmark solutions are tabulated in Table I in the Appendix.

A. Helmholtz integral method

For the case of an isovelocity halfspace and a single frequency pressure field, together with the Dirichlet (total pressure is equal to zero) boundary condition, Thorsos[17,19] has proposed a benchmark model. Hastings *et al.*,[20] Rosenberg,[8] and Stephen[21] have also used the Thorsos model to verify the application of their methods to acoustic scattering from rough sea surfaces.

The Thorsos model calculates the scattering of a finite width beam by a rough sea surface in a 2-D plane. The Helmholtz integral formulation is used to derive a surface integral equation of the first kind. An initial Gaussian shaded field, defined at zero range as a function of depth (and also used as the PE starting field), is propagated to a 1-D rough sea surface to provide the incident surface pressure field. This is used as input into the integral equation method. The solution of the surface integral equation provides the normal derivative of the total pressure field on the surface, from which the scattered field at any point can be obtained. The total acoustic field is obtained as the sum of the scattered and incident pressure fields.

While the Helmholtz method applies strictly to 2-D scattering, the PE method is derived from a cylindrically sym-

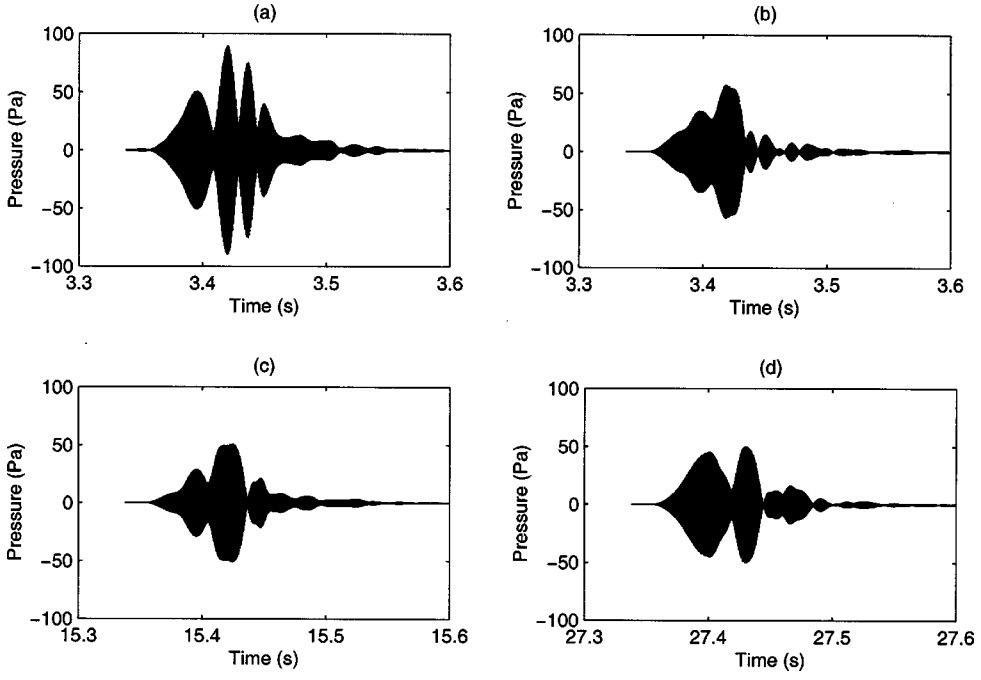


FIG. 11. Received pulses in a winter environment at a receiver depth of 75 m; (a) flat surface, (b) rough sea surface at $t=0$ s, (c) rough sea surface at $t=12$ s, and (d) rough sea surface at $t=24$ s.

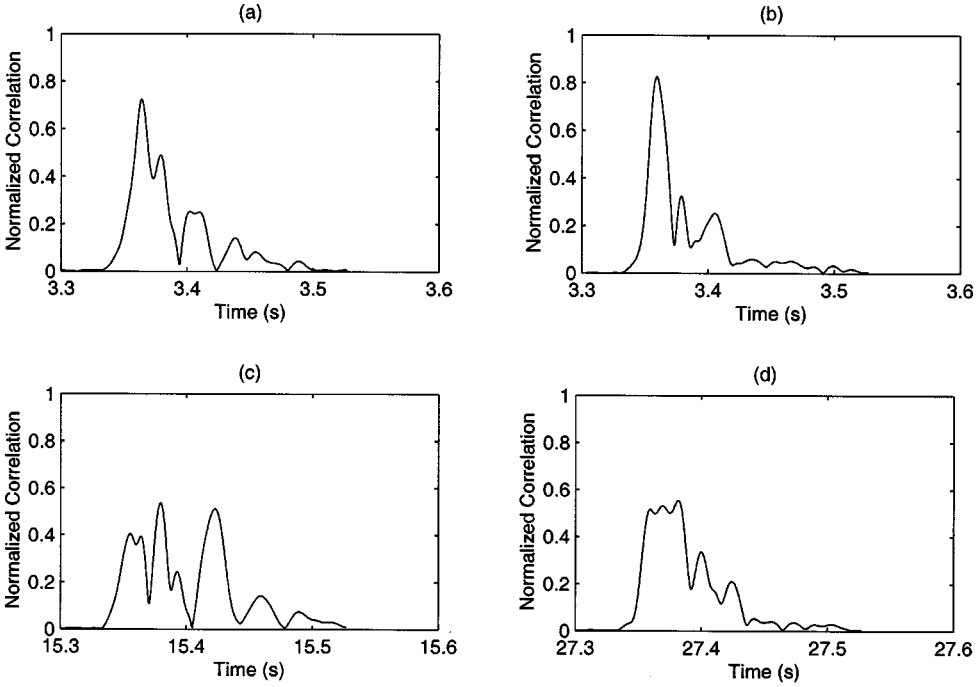


FIG. 12. Envelope of the replica correlator output in a summer environment at a receiver depth of 10 m; (a) flat surface, (b) rough sea surface at $t=0$ s, (c) rough sea surface at $t=12$ s, and (d) rough sea surface at $t=24$ s.

metric 3-D problem. Therefore, to compare the two methods it is necessary to remove the factor of $r^{-1/2}$ in the RAM(v) solutions to eliminate cylindrical spreading.

B. Comparisons to reference solutions

1. Thorsos test case 1

This test case is taken from Thorsos[19] and consists of an incident beam at an elevation angle of 10° in an isovelocity halfspace. The sound speed in the water is 1500 ms^{-1}. The source is at a depth of 66.12 m with a Gaussian width of 27.55 m. The solution is computed at a range of 750 m along a vertical profile in the water column (to a depth of 500 m) and at a frequency of 400 Hz for the Pierson–Moskowitz spectrum with a wind speed of 15 ms^{-1}. Figure 1 illustrates the pressure magnitude and phase difference versus depth obtained from the Helmholtz method (dashed line) and RAM(v) (solid line) and highlights the excellent agreement between the two methods.

2. Pulse test case

In Sec. V propagation of a linear frequency modulated pulse with center frequency 1500 Hz is modeled. It is therefore instructive to compare RAM(v) with the Helmholtz method at this center frequency.

For this test case the environment again consists of an isovelocity halfspace with sound speed equal to 1500 ms^{-1}

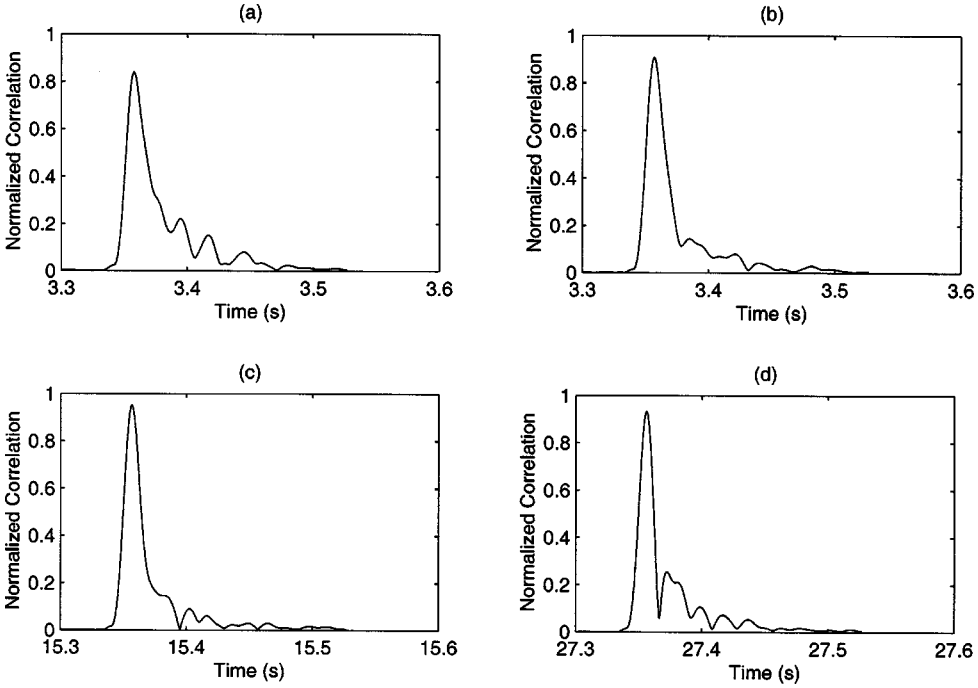


FIG. 13. Envelope of the replica correlator output in a summer environment at a receiver depth of 35 m; (a) flat surface, (b) rough sea surface at $t=0$ s, (c) rough sea surface at $t=12$ s, and (d) rough sea surface at $t=24$ s.

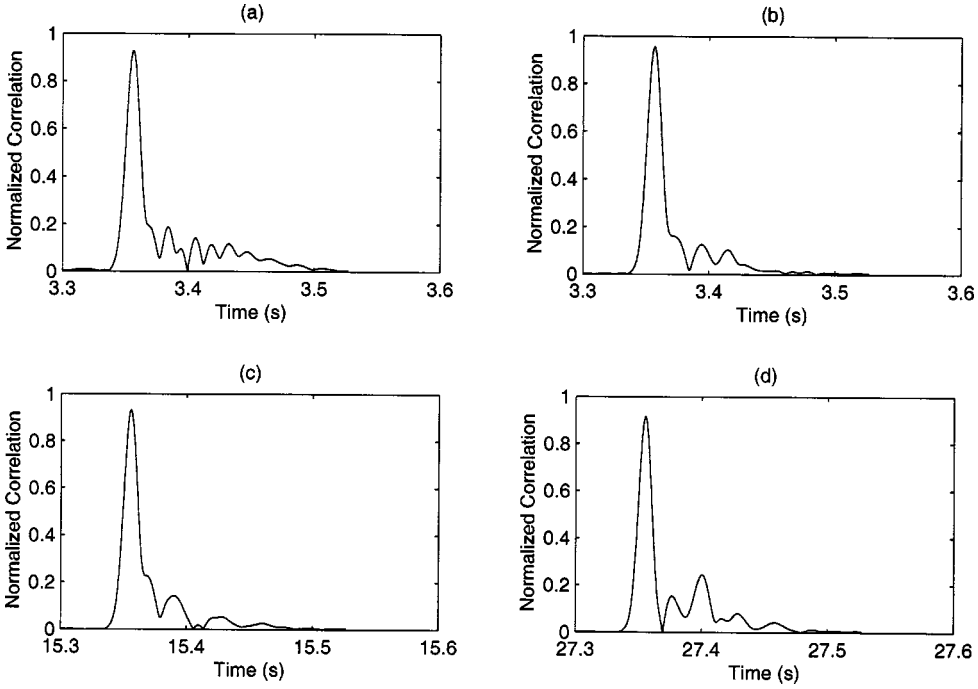


FIG. 14. Envelope of the replica correlator output in a summer environment at a receiver depth of 75 m; (a) flat surface, (b) rough sea surface at $t=0$ s, (c) rough sea surface at $t=12$ s, and (d) rough sea surface at $t=24$ s.

but with a source depth of 75 m and a Gaussian width of 2.686 m. The source frequency is 1500 Hz. The solution is computed at a range of 450 m, through the water column (to a depth of 500 m) for an incident beam at an elevation angle of 30° using the Pierson-Moskowitz spectrum with a wind speed of 15 ms^{-1}.

Figures 2 and 3 illustrate the pressure magnitude and phase difference versus depth for the Helmholtz method (dashed line) and the RAM(v) output (solid line) for a flat sea surface and the rough sea surface, respectively. It is clear that the surface scattering implementation can provide good results for shallow incident and scattering angles, but agreement decreases at steeper scattering angles. The method is therefore applicable to typical long range propagation problems in which propagation is negligible at angles beyond the water-seafloor critical angle. Critical angles are typically less than about 30° for most practical ocean-acoustic problems.

The close agreement between the RAM(v) and the Helmholtz solutions for a flat sea surface (Fig. 2) indicates that the wide angle approximation used within RAM(v) is capable of modeling propagation at steep angles. The reason, therefore, for the poor performance at steep scattering angles (Fig. 3) is due to the vertical interfaces of the stair-step approximation of the rough sea surface. Solving for the modified depth operator and pressure field within RAM(v)

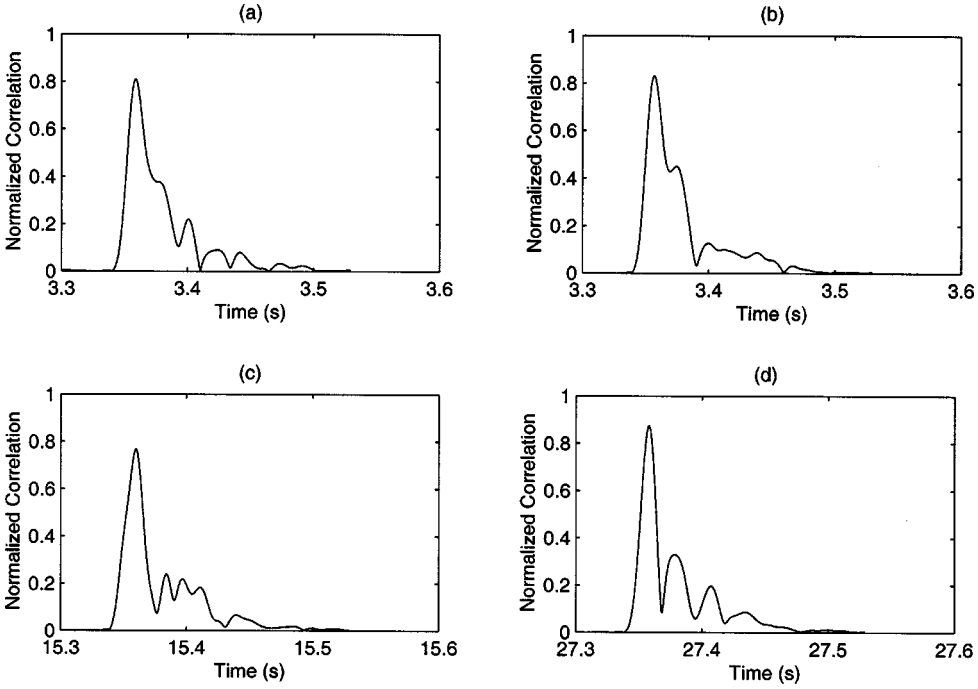


FIG. 15. Envelope of the replica correlator output in a winter environment at a receiver depth of 10 m; (a) flat surface, (b) rough sea surface at $t=0$ s, (c) rough sea surface at $t=12$ s, and (d) rough sea surface at $t=24$ s.

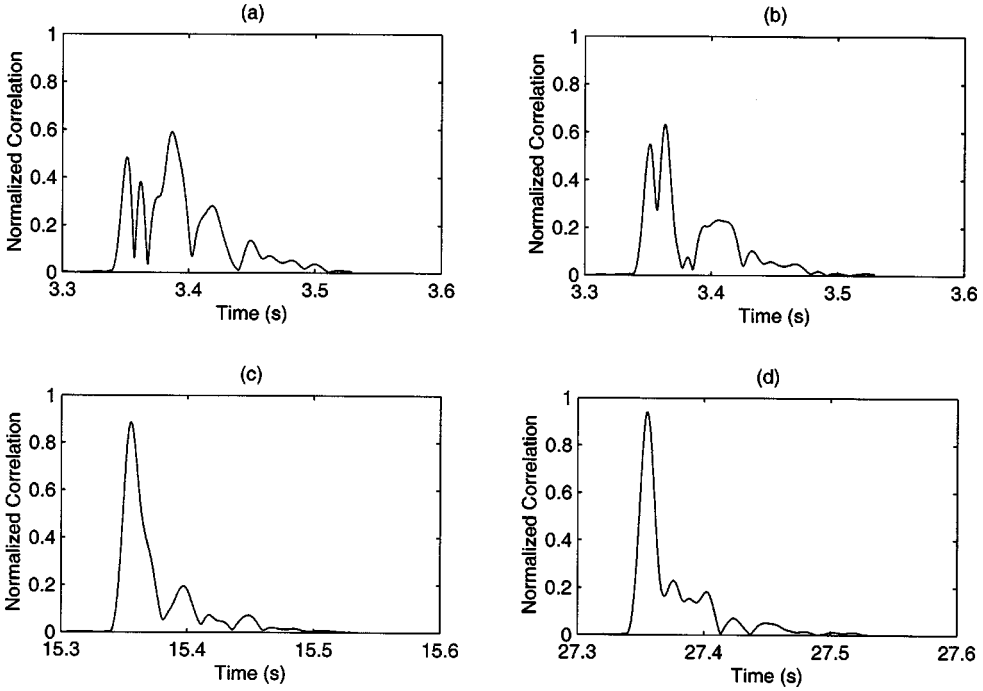


FIG. 16. Envelope of the replica correlator output in a winter environment at a receiver depth of 35 m; (a) flat surface, (b) rough sea surface at $t=0$ s, (c) rough sea surface at $t=12$ s, and (d) rough sea surface at $t=24$ s.

only conserves energy across the vertical interfaces for propagation close to the horizontal.[9] Hence for higher scattering angles, the modelling of the scattering process is less accurate.

Comparison of Fig. 2(a) with Fig. 3(a) highlights the dramatic redistribution of energy with angle that the rough surface produces. This is further illustrated in Fig. 4, in which the propagation loss corresponding to Fig. 3 is plotted as a function of depth and range.

3. Comparison to OASES

In this section we present a set of comparisons between OASES[22] and RAM(v) for a Gaussian sea surface spectral density. OASES is a general purpose code for modeling seismo-acoustic propagation in horizontally stratified waveguides using wave number integration. It computes the coherently averaged pressure field for a Gaussian sea surface spectrum, using perturbation theory. To provide the comparison it is necessary to generate the received pressures using RAM(v) from an ensemble of sea surface realizations. The resulting pressures are coherently averaged and compared with the output of OASES.

The environment for these comparisons is defined in Table II in the Appendix. The source is omni-directional and at a depth of 15 m. A rms height of 0.534 m and a correlation length of 100 m are used to generate the ensemble of

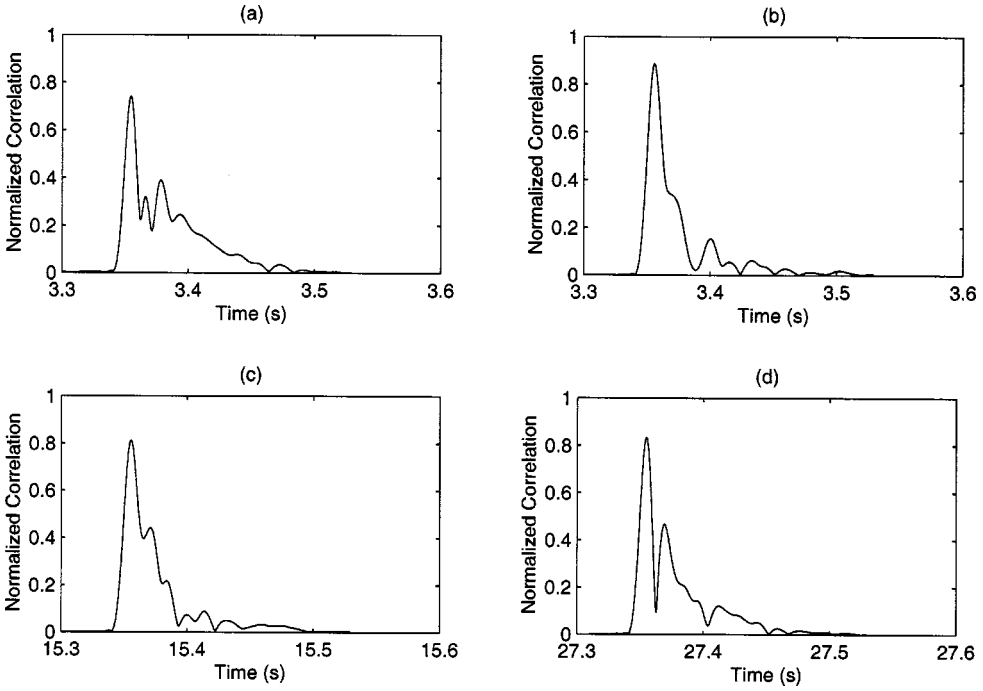


FIG. 17. Envelope of the replica correlator output in a winter environment at a receiver depth of 75 m; (a) flat surface, (b) rough sea surface at $t=0$ s, (c) rough sea surface at $t=12$ s, and (d) rough sea surface at $t=24$ s.

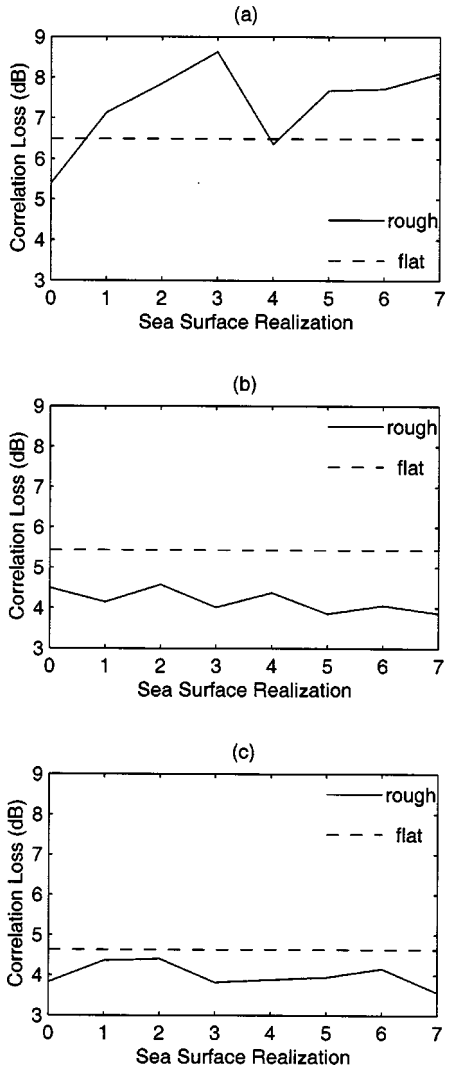


FIG. 18. Correlation loss for a summer environment at three receiver depths; (a) 10 m, (b) 35 m, and (c) 75 m.

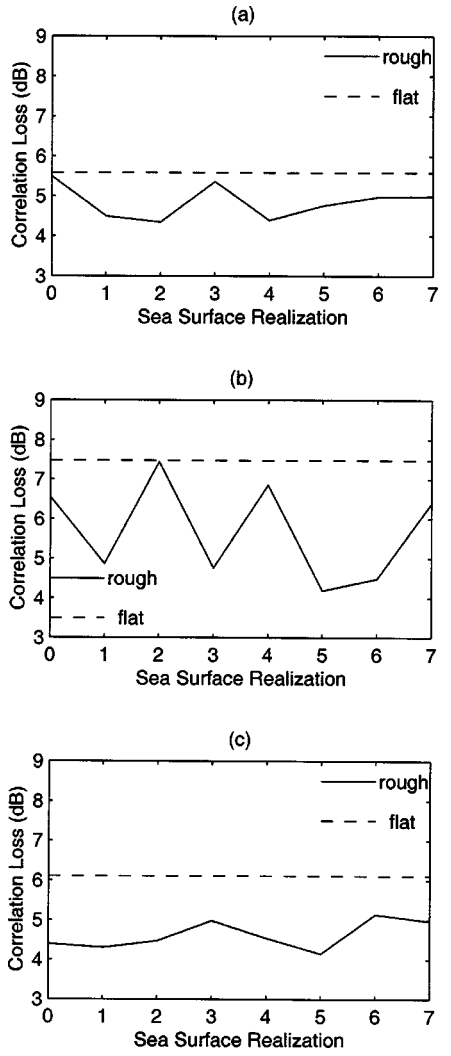


FIG. 19. Correlation loss for a winter environment at three receiver depths; (a) 10 m, (b) 35 m, and (c) 75 m.

Gaussian surfaces for the 3-D cylindrically symmetric problem. The reader is referred to Table I in the Appendix for the relevant RAM(v) convergence parameters and CPU times.

Figures 5(a) and (b) depict the propagation loss versus range for a frequency of 400 Hz for both OASES (dashed line) and RAM(v) results (solid line) for a receiver depth of 80 m for a flat surface and a set of rough surfaces, respectively. For the rough surface case the result from RAM(v) is an average computed from an ensemble of 50 realizations of the sea surface. The results at this depth are in excellent agreement. Similar agreement is found at depths of 20 and 100 m.

V. PULSE VARIABILITY IN A SHALLOW WATER WAVEGUIDE

An area of interest in sonar performance modeling is the changes in pulse characteristics between different pulse transmissions. Movement of the sea surface will clearly play a large role in causing this variability. In the present calculations the variability is modeled by creating an initial sea surface profile, performing the pulse computation and then advancing the sea surface [using the dispersion relation defined in Eq. (8)] and performing a new pulse computation. The sea surface is assumed to be stationary during each separate pulse transmission.

A range independent shallow water waveguide with summer and winter sound speed profiles is considered (see Table II in the Appendix for details) for the 3-D cylindrically symmetric problem. The sea surface is generated using a Pierson-Moskowitz spectral density with a wind speed of 15 ms^{-1}.

A point source at a depth of 50 m with receivers at depths 10, 35, and 75 m, each with a range of 5 km, is considered. The source transmits a set of eight 0.1-s Hanning shaded linear frequency-modulated pulses at a center frequency of 1500 Hz and swept bandwidth of 100 Hz. The peak power of each pulse is 220 dB *re* 1 μPa at 1 m. A pulse is transmitted every 4 s.

The reader is referred to Table I in the Appendix for the relevant RAM(v) convergence parameters and CPU times. To synthesise the pulses $N=190$ frequencies are used with frequency increment (or ''frequency resolution'' as given in Table I) $\Delta f = B/N$, where $B=556.64$ Hz is the full bandwidth modeled. Zero padding is used to increase the FFT

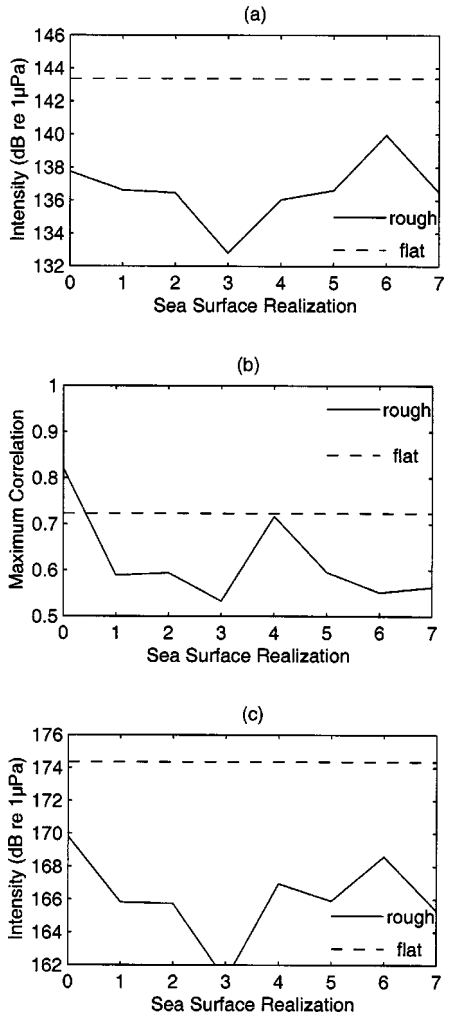


FIG. 20. Received signal properties in a summer environment at a receiver depth of 10 m; (a) intensity before replica correlation, (b) maximum normalized response out of replica correlator, and (c) intensity after replica correlation.

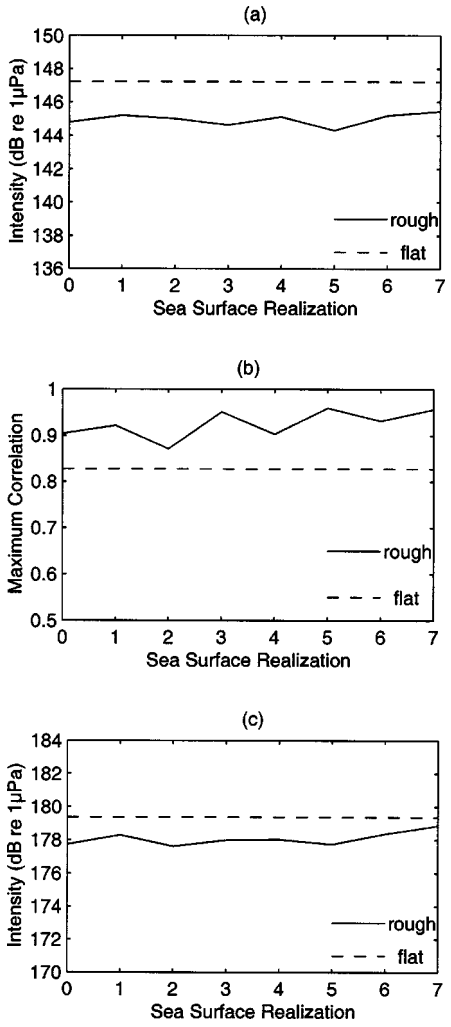


FIG. 21. Received signal properties in a summer environment at a receiver depth of 35 m; (a) intensity before replica correlation, (b) maximum normalized response out of replica correlator, and (c) intensity after replica correlation.

size to 2^{13} which leads to a sampling frequency of 24 kHz for the time domain waveforms.

Figures 6–8 illustrate the received pulses for the summer environment for the flat surface case and the rough sea surfaces cases at $t=0$ s, $t=12$ s, and $t=24$ s for each receiver depth. Figures 9–11 are the corresponding plots for the winter environment example.

It is evident, from Figs. 6–11, that there are considerable variations between the level and structure of the received pulses. In particular, for the summer environment the propagation paths reaching the 10-m receiver interact more strongly with the sea surface than the paths that reach the deeper receivers. As a consequence, there is much more pulse variability measured at the 10-m receiver.

To quantify the important features of the received pulses and their effect on a sonar system it is useful to calculate the replica correlation function. For simplicity, replica correlation of the transmit pulse is performed directly with each of the received pulses.

Figures 12–14 and Figs. 15–17 illustrate the envelope of the replica correlator for the flat surface case and the rough sea surface cases at $t=0$ s, $t=12$ s, and $t=24$ s for each receiver depth for the summer and winter environments, respectively.

For the summer environment example there is considerable variability of the correlation function at the 10-m receiver depth (Fig. 12), which is in contrast to the responses at the 35- and 75-m receiver depths (Figs. 13 and 14). The compactness of the correlation function can be quantified by the correlation loss (CL) defined in Eq. (9).

$$CL = -10\times\log_{10}\left\{\frac{\int_{t_0-(1/2)B}^{t_0+(1/2)B}|r(t)|\,dt}{\int_{-\infty}^{\infty}|r(t)|\,dt}\right\}, \tag{9}$$

where t_0 is the time at which the correlator is a maximum, B is the swept bandwidth of the replica, and $r(t)$ is the correlation function. Figures 18 and 19 illustrate the correlation loss at each receiver depth for both the summer and winter environments. A dashed line represents the value for a flat sea surface and a solid line represents values for each rough sea surface realization. The eight sea surface realizations are labeled from 0 to 7.

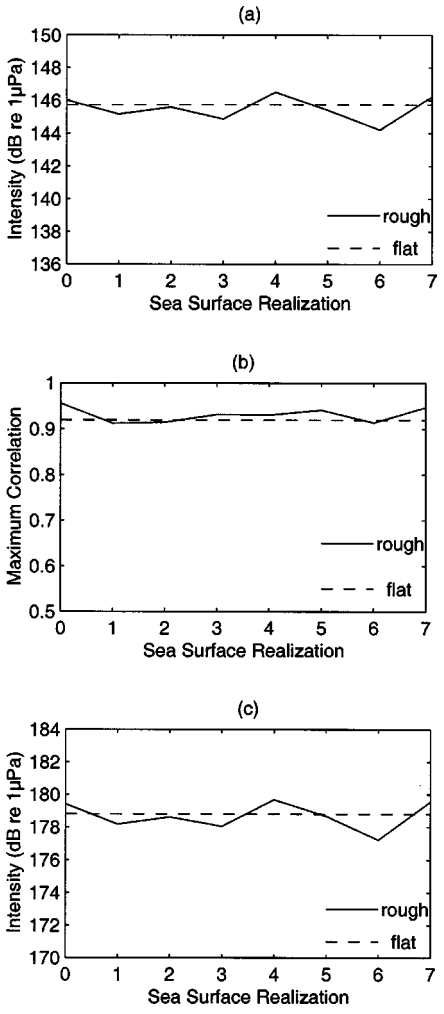


FIG. 22. Received signal properties in a summer environment at a receiver depth of 75 m; (a) intensity before replica correlation, (b) maximum normalized response out of replica correlator, and (c) intensity after replica correlation.

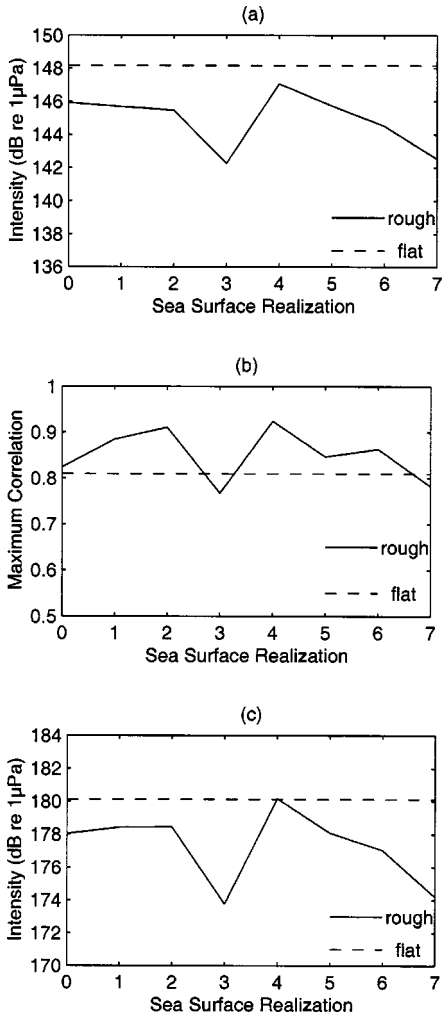


FIG. 23. Received signal properties in a winter environment at a receiver depth of 10 m; (a) intensity before replica correlation, (b) maximum normalized response out of replica correlator, and (c) intensity after replica correlation.

The reduction in the correlation loss for the rough sea surface cases compared to the flat surface cases is due to the reduction of the later multipath arrivals in the received pulses. This reduction is a consequence of the scattering of these arrival paths from the rough sea surface into steeper angle paths. These steeper angle paths suffer greater losses when interacting with the seafloor. The early arrivals are also scattered, but since these paths graze the sea surface at shallower angles they are scattered to a lesser degree.

For each of the nine sea surfaces the maximum correlation between the replica and received pulse is found. At the moment of maximum correlation the intensities before and after the correlation process are also determined. These intensities, along with the corresponding maximum correlation, are illustrated in Figs. 20–22 for each receiver depth for the summer environment. A dashed line represents the value for a flat sea surface and a solid line represents values for each of the eight rough sea surface realizations. Figures 23–25 show the corresponding plots for the winter environment.

Equation (10) describes the relationship between the intensity before and after replica correlation,

$$I_{\text{out}} = I_{\text{in}} + 10\log_{10}(N) + 10\log(\rho^2), \tag{10}$$

where $I_{\text{out}}(\text{dB}\,re\ 1\ \mu\text{Pa})$ is the intensity after replica correlation, $I_{\text{in}}(\text{dB}\,re\ 1\ \mu\text{Pa})$ is the intensity before replica correlation, ρ is the maximum correlation, and N is the number of samples in the replica correlation process (in this case $N = 2400$).

The maximum correlation is a measure of the similarity of the receive pulse with the replica and is related to the gain of the correlation process. For an undistorted receive pulse the maximum correlation is 1 and the gain of the intensity is given by $10\log_{10}(N)$.

It can be seen that for the summer environment, and at a receiver depth of 10 m, the maximum correlation of the eight sea surface realizations are, in general, less than the flat surface case. In the flat sea surface case the maximum correlation is 0.72, which corresponds to a 2.9-dB loss in gain as compared to an undistorted receive pulse.

For the other cases, the maximum correlation is generally higher for a rough surface. This highlights the fact that the presence of a rough sea surface can reduce the pulse distortion of the receive signal. However, the received inten-

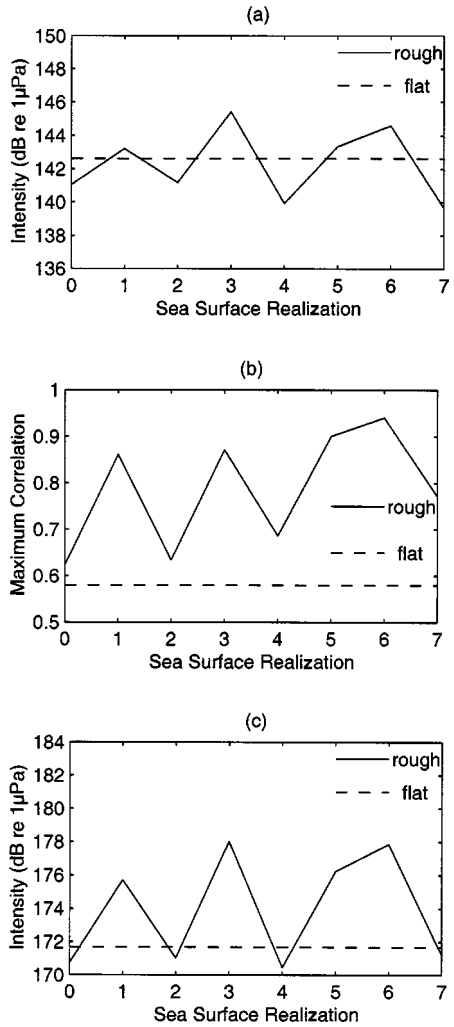


FIG. 24. Received signal properties in a winter environment at a receiver depth of 35 m; (a) intensity before replica correlation, (b) maximum normalized response out of replica correlator, and (c) intensity after replica correlation.

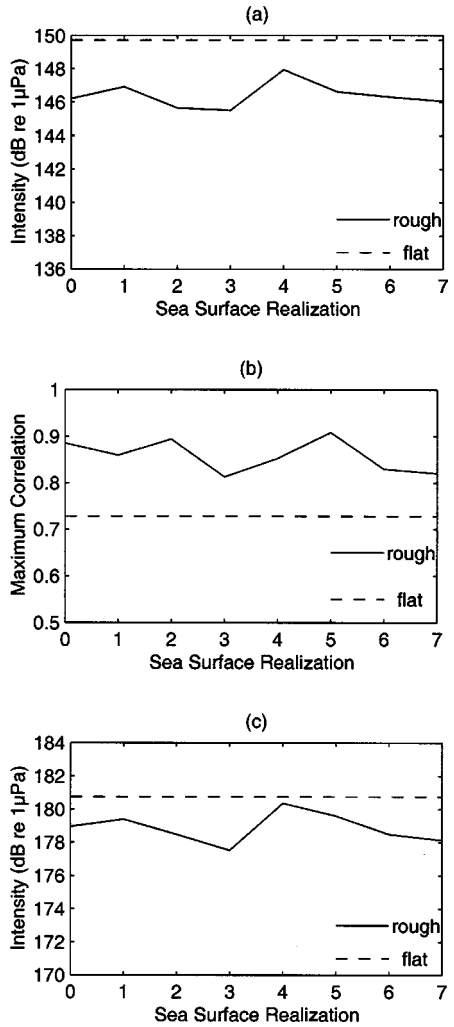


FIG. 25. Received signal properties in a winter environment at a receiver depth of 75 m; (a) intensity before replica correlation, (b) maximum normalized response out of replica correlator, and (c) intensity after replica correlation.

sities of the flat surface cases are, in general, larger than the rough sea surface cases and the postprocessing intensity (I_{out}) of a flat surface tends to be higher than for a rough sea surface. Thus, it is evident that there is a trade-off between the lower propagation loss and higher pulse distortion of the flat sea surface cases as opposed to the higher propagation loss and lower pulse distortion of the rough sea surface cases.

Figures 12–17 also highlight the sensitivity of the position in time of the maximum correlation to the sea surface boundary. This has a corresponding effect on the estimated range of the source. If the received pulse is assumed to travel at the depth average sound speed of the water column and the maximum correlation time is taken to be the arrival time of the pulse, the corresponding range can be estimated. Figures 26 and 27 give the estimated range for each of the received pulses for the summer and winter environments respectively. In general, the estimated ranges for both the rough and flat sea surface cases are within 10 m of the actual range. However, for the 10-m receiver in the summer environment the range estimate can be up to 45 m in error. It is also worth noting that for the winter environment, with a flat sea surface for the 35-m receiver depth, the estimated range is in error by 50 m.

VI. CONCLUSIONS

It has been demonstrated that a rough sea surface boundary can be represented as a range-dependent internal interface within RAM(v). Comparisons to reference solutions of a Helmholtz integral equation illustrate this method is very accurate for shallow scattering angles (which are the most important for long range propagation). The implementation of a variable computational depth step scheme within RAM(v) enables a more computationally efficient treatment of interfaces between different environmental layers.

The pulse variability due to a moving sea surface in a shallow water waveguide, as modeled by SPUR, highlights the significant effect of a rough sea surface boundary on the received pulse structure. It is interesting to observe that a rough sea surface can reduce the correlation losses associated with geometric pulse distortion. The trade-off between the lower propagation loss and higher pulse distortion of a flat-sea surface and the higher propagation loss and lower pulse distortion of a rough sea surface is apparent. It is also inter-

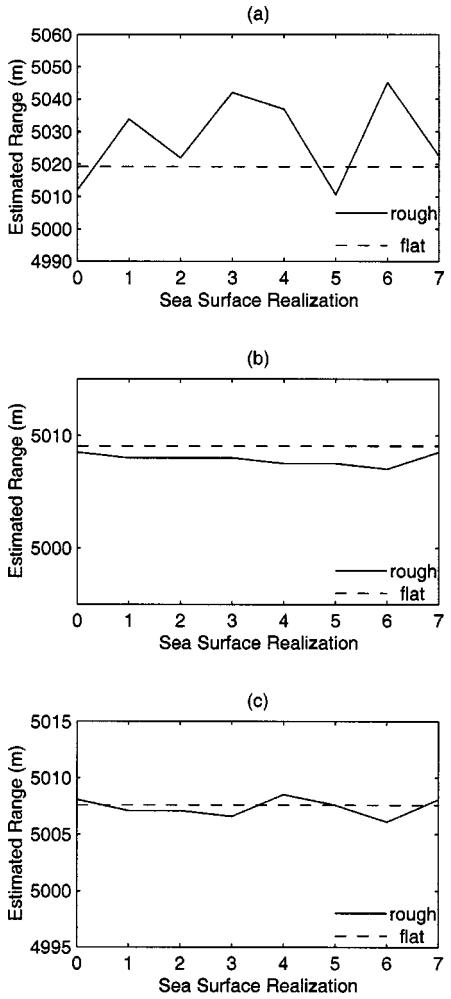


FIG. 26. Range errors for the summer environment at three receiver depths; (a) 10 m, (b) 35 m, and (c) 75 m.

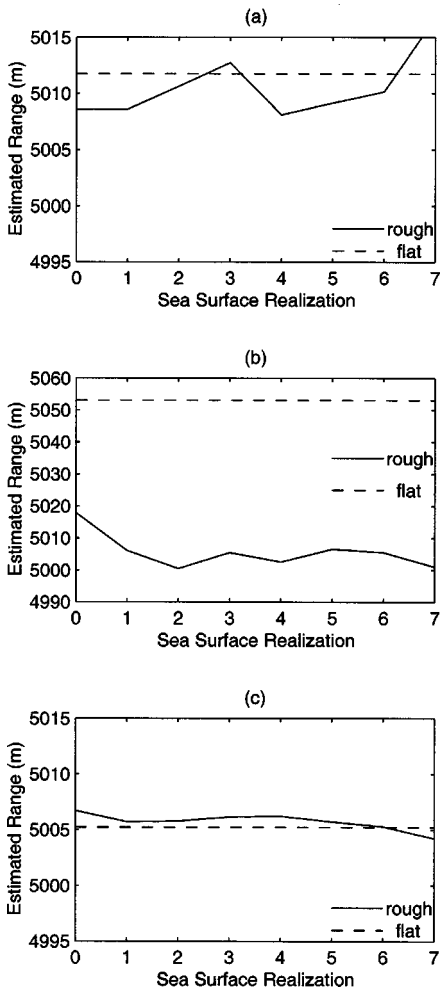


FIG. 27. Range errors for the winter environment at three receiver depths; (a) 10 m, (b) 35 m, and (c) 75 m.

esting to note that surface duct propagation with a rough sea surface can significantly change the estimated range of a target.

By relaxing the assumption that the sea surface is stationary during the travel time of the propagating pulse the method presented in this paper can be extended to include Doppler effects arising from a moving sea surface boundary.

ACKNOWLEDGMENTS

This research was funded by the United Kingdom Ministry of Defense Applied Research Program. The authors would like to thank M. Collins for helpful suggestions on the variable depth step implementation.

APPENDIX: MODELING PARAMETERS

TABLE I. RAM(v) convergence and Fourier synthesis parameters together with corresponding CPU times.

Scenario	dr[a] (λ)	dz[b] (λ)	np[c]	f_0[d] (Hz)	Frequency (Hz)	f_s[e] (kHz)	Range (km)	Depth (km)	CPU time[f] (h)
Thorsos test case 1	0.1	0.02	8	⋯	400	⋯	0.750	1300	0.13
LFM pulse test case	0.05	0.006 25	8	⋯	1500	⋯	0.450	750	4.46
OASES	1	0.04	8	⋯	400	⋯	60.0	340	0.12
Waveguide	0.4	0.05	8	2.93	1221.68–1778.32	24	5.0	120	29.5

[a] dr is the computational range step.
[b] dz is the coarse computational depth step (away from the interfaces).
[c] np is the number of terms in the Padé series used within RAM(v).
[d] f_0 is the frequency resolution (for pulse calculations).
[e] f_s is the sampling frequency (for pulse calculations).
[f] Pentium II, 333 MHz, 512 MB.

TABLE II. Water column and sediment parameters of each modeled environment.

Environment	H_w[a] (m)	z[b] (m)	c[c] (ms^{-1})	c_{sed}[d] (ms^{-1})	ρ_{sed}[e] (kgm^{-3})	α_{sed}[f] ($dB\lambda^{-1}$)
Thorsos test case 1	⋯		1500	⋯	⋯	⋯
LFM pulse test case	⋯		1500	⋯	⋯	⋯
OASES	100	0	1493.29	1631	1800	0.49
		100	1495.34			
Waveguide: summer	100	0	1494	1560	1800	1.00
		20	1494.3			
		50	1490.6			
		100	1491.3			
Waveguide: winter	100	0	1490	1560	1800	1.00
		100	1492			

[a]H_w is the water depth.
[b]z is the depth.
[c]c is the sound speed in the water column (with linear interpolation between the numbers given).
[d]c_{sed} is the sound speed in the sediment.
[e]ρ_{sed} is the density in the sediment.
[f]α_{sed} is the attenuation in the sediment.

[1] R. L. Holford, "Scattering of sound waves at the ocean surface: A diffraction theory," J. Acoust. Soc. Am. **70**, 1103–1115 (1981).
[2] R. L. Holford, "Scattering of sound waves at a periodic, pressure-release surface: An exact solution," J. Acoust. Soc. Am. **70**, 1116–1128 (1981).
[3] G. V. Norton and J. C. Novarini, "Enhancement of the total acoustic field due to the coupling effects from a rough sea surface and a bubble layer," J. Acoust. Soc. Am. **103**, 1836–1843 (1998).
[4] L. Fortuin, "The wave equation in a medium with a time-dependent boundary," J. Acoust. Soc. Am. **53**, 1683–1685 (1972).
[5] M. A. Leontovich and V. A. Fock, "Solutions of the problem of propagation of electromagnetic waves along the earth's surface by method of the parabolic equation," J. Exp. Theor. Phys. **16**, 557–573 (1946).
[6] L. B. Dozier, "PERUSE: A numerical treatment of rough surface scattering for the parabolic wave equation," J. Acoust. Soc. Am. **75**, 1415–1432 (1984).
[7] M. D. Collins, "Application and time-domain solution of higher-order parabolic equations in underwater acoustics," J. Acoust. Soc. Am. **86**, 1097–1102 (1989).
[8] A. P. Rosenberg, "A new rough surface parabolic equation program for computing low-frequency acoustic forward scattering from the ocean surface," J. Acoust. Soc. Am. **105**, 144–153 (1999).
[9] M. D. Collins, "User's Guide for RAM Versions 1.0 and 1.0p," Naval Research Laboratory, Washington, DC 20375 (1995).
[10] M. D. Collins, "A split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[11] M. D. Collins, "A self-starter for the parabolic equation method," J. Acoust. Soc. Am. **92**, 2069–2074 (1992).
[12] K. E. Gilbert, "Application of the parabolic equation to sound propagation in a refracting atmosphere," J. Acoust. Soc. Am. **85**, 630–637 (1989).
[13] P. A. Bello, "Characterization of randomly time-variant linear channels," IEEE Trans. Commun. Syst. **CS-11**, 360–393 (1963).
[14] K. Aki and P. G. Richards, *Quantitative Seismology: Theory and Methods* (Freeman, 1980), Vol. I, p. 174.
[15] R. S. Kulkarni, W. L. Seigmann, M. D. Collins, and B. E. McDonald, "Nonlinear pulse propagation in shallow-water environments with attenuating and dispersive sediments," J. Acoust. Soc. Am. **104**, 1356–1362 (1998).
[16] E. I. Thorsos, "Acoustic scattering from a Pierson-Moskowitz sea surface," J. Acoust. Soc. Am. **88**, 335–349 (1990).
[17] E. I. Thorsos, "The validity of the Kirchoff approximation for rough surface scattering using a Gaussian roughness spectrum," J. Acoust. Soc. Am. **83**, 78–92 (1988).
[18] W. J. Pierson, Jr. and L. Moskowitz, "A proposed spectral form for fully developed wind seas based on the similarity theory of S. A Kitaigorodski," J. Geophys. Res. **69**, 5181–5190 (1964).
[19] E. I. Thorsos, "Test Case 1: Sea surface forward scattering," in *Benchmark Solutions in Reverberation and Scattering: Proceedings of the Reverberation and Scattering Workshop*, edited by D. B. King, S. A. Chin-Bing, J. A. Davis, and R. B. Evans, Naval Research Laboratory Book Contribution NRL/BE/7181-96-001 (U.S. Government Printing Office, Washington, DC, 1994), pp. 3.2–3.20.
[20] F. D. Hastings, J. B. Schneider, and S. L. Broschat, "A finite-difference time-domain solution to scattering from a rough pressure-release surface," J. Acoust. Soc. Am. **102**, 3394–3400 (1997).
[21] R. A. Stephen, "Modeling sea surface scattering by the time-domain finite-difference method," J. Acoust. Soc. Am. **100**, 2070–2078 (1996).
[22] H. Schmidt, OASES User Guide, Version 2.1, 1997.

At-sea measurements of sound penetration into sediments using a buried vertical synthetic array

Harry J. Simpson and Brian H. Houston
Physical Acoustics Branch Code 7136, Naval Research Laboratory, Washington, D.C. 20375

Steve W. Liskey, Philip A. Frank, Alain R. Berdoz, and Larry A. Kraus
SFA, Inc., Largo, Maryland 20774

Carl K. Frederickson
Physics and Astronomy, University of Central Arkansas, Conway, Arkansas 72035

Steve Stanic
Acoustic Simulations, Measurements and Tactics Code 7180, Naval Research Laboratory, Stennis Space Center, Mississippi 39529

(Received 26 April 2002; revised 23 May 2003; accepted 30 May 2003)

Acoustic bottom penetration experiments were carried out in a medium-grain sandy bottom at a site in St. Andrews Bay, Florida. These investigations used a new buried, vertical, one-dimensional synthetic array system where a small hydrophone was water-jetted into the sediment to a depth of ~2 m. Once buried, this hydrophone was mounted to a vertical robotics stage that translated the hydrophone upward in 1-cm increments. A broadband (3 to 80 kHz) spherical source, positioned 50 cm above the sediment–water interface, was used to insonify the sediment. Measurements were made with insonification angles above and below the critical angle by changing the horizontal distance of the source relative to the insertion point. This new measurement system is detailed, and results are presented that include temporal, frequency, and wavenumber analysis for natural and roughened interfaces. The measured compressional sound speed and attenuation are shown to be self-consistent using the Kramers–Kronig relation. Furthermore, only a single fast compressional wave was observed. There was no observation of a second slower compressional wave as predicted by some applications of the Biot model to unconsolidated water-saturated porous media. [DOI: 10.1121/1.1594192]

PACS numbers: 43.20.Jr, 43.30.Hw [RAS]

I. INTRODUCTION

The sandy bottom of littoral environments is a rich and complex acoustic medium. Due to the complex biological, tidal, and storm forces, these bottoms are constantly changing and moving. Thus, characterizing the acoustic properties is difficult. The work reported here is inspired, in part, by the need for high-quality sound-speed and attenuation measurements in these types of bottoms. Towards this end, a method is employed to spatially map the acoustic field in marine sediments yielding a high enough resolution to clearly delineate the different waves propagating in the bottom and to more directly measure their sound speeds and attenuations.

Typically, water-saturated sandy environments have sound speeds greater than the water column, resulting in a critical angle below which refractive energy does not propagate.[1,2] Such below-critical angle insonifications are of most interest due to implications with regard to long-range buried target detection. Several mechanisms have been proposed that suggest possible subcritical penetration mechanisms, but better experimental techniques are still required in order to examine this issue. We report here on a new, one-dimensional synthetic buried array technique that can provide high-resolution broadband measurements of the sound waves propagating in littoral environments. These measurements were designed to accurately measure the sound speeds and attenuations, and to provide new insight into the mechanisms responsible for energy entering the bottom for below-critical angle insonifications.

There are several theories that provide mechanisms for getting energy into marine sediments for the case of below the critical angle insonification. The first is associated with the evanescent wavefield that exists for below-critical angle insonifications.[3,4] Although the wavefield decays exponentially with distance into the bottom, this energy can be scattered from inhomogeneities in the bottom near the interface. The second mechanism, proposed by Chotiros,[5] predicts a slower, second compressional wave based on modeling the water-saturated sandy bottom as a Biot medium. Here, no critical angle exists for the second slower wave. Another mechanism has been proposed that is associated with the scattering of acoustic energy from volume inhomogeneities[6,7] and interface roughness.[2,8–10]

Shear wave energy is also a viable mechanism for energy penetration into sea bottoms for below-critical angle insonifications.[11,12] However, two conditions must be present for the shear wave to provide substantial energy in the bottom. The first is the shear wave speed must be close to the compressional wave speed.[11,12] For the bottom investigated during these measurements, and for most bottoms of interest where objects scour into the sediment, the shear wave speed is typically much lower than the compressional wave speed.[13] For this type of sandy bottom the shear wave speed has been measured between 50 and 150 m/s.[13] Also, the shear wave is associated with very low frequencies. Previous measurements have reported shear wave energy effects in hard bottoms for frequencies below 500 Hz, but little shear

wave energy effects in soft bottoms or for frequencies above 500 Hz.[12] The measurements reported here are for soft bottoms and for the frequency range between 3 and 80 kHz.

To better understand these mechanisms, several laboratory and littoral environment measurements have been conducted in recent years. Chotiros[5] conducted a laboratory measurement that supports the concept of a second slower compressional wave predicted by application of Biot's theory to water-saturated sand. However, Simpson and Houston[2] conducted sound-speed measurements in a separate set of laboratory experiments using a two-dimensional synthetic array technique that does not show the existence of a second slower compressional wave. Although these laboratory measurements gave some insight into understanding these sandy bottoms, high-resolution synthetic array techniques have yet to be applied in a littoral environment.

To measure the sound propagating in the ocean bottoms, hydrophones are buried to form one-, two-, or three-dimensional arrays.[1,5] Sophisticated techniques can be used to reconstruct the wavefronts and infer sound speeds and attenuations. However, most of these measurements, reported to date, lack the spatial resolution and relative spatial accuracy in alignment to directly measure the sound speeds and attenuations. One approach that can be used to directly measure these physical parameters involves mounting receivers on a rigid structure that is then buried. This technique can provide a tremendous amount of information, but the hydrophone system must be designed with great care. Once assembled, the spacing and length are set, limiting the lower and upper frequency range. Also, due to the number of receivers, a large data acquisition system is required to simultaneously measure all the hydrophones. Since different hydrophones are used, each receiver has to be calibrated; the calibration is used later in postprocessing of the data to normalize the measured amplitude and phase. Such a process is very sensitive to noise and must be done with great care. Another inherent problem with such a hydrophone array is the scattering of sound from each hydrophone that is then measured by the other array elements.

Simpson and Houston[2] demonstrated the use of a synthetic array in the laboratory for measuring the sound speeds and attenuations of acoustical waves in a water-saturated sandy bottom that has been shown to work well in this setting. One of the objectives of the work reported here is to verify the applicability of the synthetic array technique in an at-sea environment. This technique is best suited to studying coherent processes and offers several advantages over a fixed array. The data acquisition system is minimized due to only one hydrophone being sampled. Since a single hydrophone is buried, the environmental disturbance is minimized as well. Additional benefits include the elimination of almost all of the calibration issues and of the mutual interference issues associated with scattering from neighboring hydrophones. Also, the number of sensor locations is programmable and can be changed with each measurement. Finally, the length of the array is only limited by the length of the positioning hardware and by how deep a sensor can be placed in the bottom. Due to the accuracy of the modern positioning encoder technology, the relative distance between each receiver position is precisely known. Since the relative spacing for these measurements is constant, the data processing and analysis is straightforward. The measurement system reported here has been designed for a broad frequency range between 3 and 80 kHz. However, the system can be easily adapted to operate at lower frequencies.

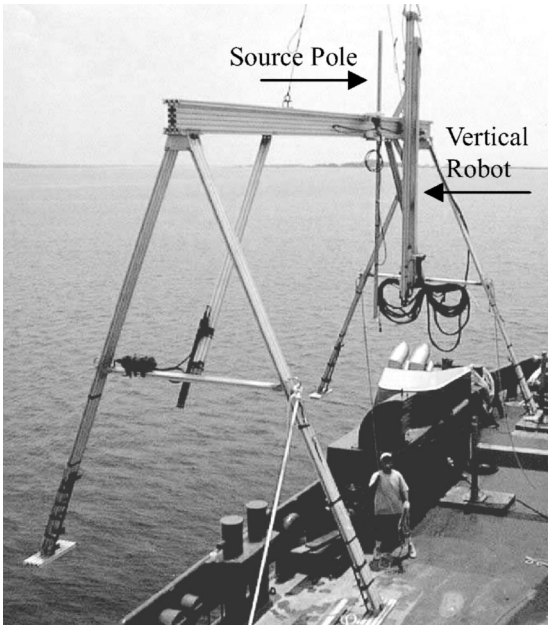


FIG. 1. A photograph of the measurement system used for the littoral measurements in St. Andrews Bay. The vertical structure to the right and extending above the backbone is the vertical robotics stage used to position the buried hydrophone. The thinner vertical pole near the center of the backbone is the source mount. The source mount is used to position the source 50 cm above the sandy bottom and slides along the backbone to move the source from directly over the insertion point to 4 m measured horizontally from the insertion point.

The buried one-dimensional synthetic array apparatus and techniques used here are based on a two-dimensional system developed for laboratory measurements at the Naval Research Laboratory.[2] A light-weight and easily deployable one-dimensional system was built and used in an experiment conducted at 30 07.898 N and 85 39.282 W in the eastern part of St. Andrews Bay, Florida. Section II will detail the measurement hardware that was designed and built to facilitate these measurements. Section III will detail the basic measurements taken in St. Andrews Bay. Finally, Sec. IV will discuss the experimental results including sound-speed analysis, attenuation analysis, wave number-frequency analysis, and the implications of the trace velocities measured by the vertical array. Also discussed are some preliminary roughened interface measurements.

II. MEASUREMENT SYSTEM

The core of the apparatus consists of a vertical robotics stage mounted to an extruded aluminum beam, as shown in Fig. 1. This 5.5 m long aluminum I-beam is mounted 3.7 m above the bottom on ''A-frame'' legs. On this member, a 3 m long vertically mounted robot is used to pull a buried hydrophone up through the bottom in 1 cm increments. The source is also mounted to the I-beam with a stainless-steel pole and is positioned 50 cm above the fluid–sediment inter-

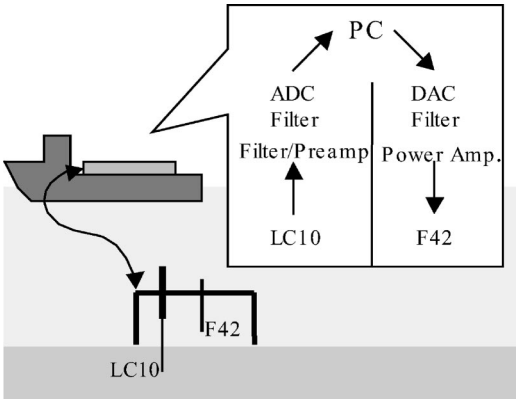


FIG. 2. A block diagram of the measurement system. The backbone, source, and receiver are deployed over the side of the ship and set on the bottom. The supporting electronics remain on the ship and are connected via cables to the submersible hardware. The onboard electronics consist of input electronics including a digital to analog arbitrary waveform generator, a 4-pole Butterworth filter, and a power amplifier that is used to drive the source. The receive electronics consist of an antialiasing filter, a preamp, a second analog filter, and a digitizer. All the electronics are controlled by a PC-based data acquisition system.

face. The source is attached to the I-beam with a sliding mount that can be positioned along the length of the I-beam, from directly over the hydrophone insertion point to a horizontal location 4 meters away. The angle, from the insertion point to the source, measured up from the sand–water interface, ranges from 90° to 7° for these measurement configurations. The critical angle for this sand–water interface is approximately 27°.

After the placement of the measurement system on the seafloor, the source is moved to the desired horizontal location and the hydrophone is buried by divers using a water jet. Typically, at each new site, the water jet is first inserted into the bottom without the hydrophone to make a pathway for the hydrophone. Care must be taken in this process to insure that the water jet system is free of gas bubbles. Once the path is cleared to approximately 2.5 m below the sand–water interface, the hydrophone is inserted with the water jet. The water jet is then removed and the sand–water interface is shaped to its original contour.

The hydrophone used for these measurements was a Celesco Industry LC-10 with dimensions of 53 mm in length and 8.6 mm in diameter. A 7 m long coaxial cable, measuring 4 mm in diameter, connects the hydrophone to the underwater preamplifier. To strain-relieve the hydrophone, the hydrophone is connected to a 1.3 mm diameter stainless-steel cable. The stainless-steel cable extends 50 cm past the hydrophone where a small fish hook is connected. This is used to pull the hydrophone down into the sediment with the water jet nozzle. The other end of the stainless-steel cable is connected to the robot and is used to incrementally pull the buried hydrophone up through the sediment.

The underwater electronics are connected with the shipboard control and data acquisition systems via cables. A single computer (PC) is used for coordinated robotics control and data acquisition, as diagramed in Fig. 2. The PC has a controller card that positions the vertical stage using a stepper motor with feedback supplied by an attached rotary re-

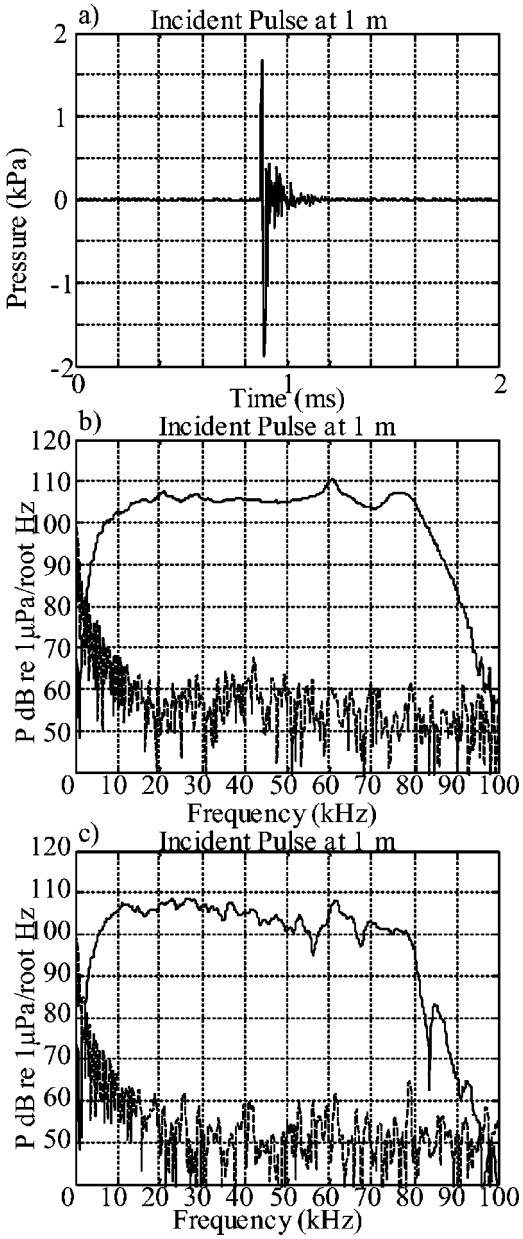


FIG. 3. The temporal and frequency response of the source–receiver system. (a) The temporal response with the source and receiver separated by 1 m in the water column. (b) The Fourier transform of (a) showing the bandwidth of the measurement. Also shown in (b) is the acousto-electrical noise floor for this measurement. (c) The frequency response of the system with the hydrophone buried 15 cm in the water-saturated sand and the source mounted 50 cm directly above the interface. Also shown in (c) is the associated noise floor for that measurement.

solver. The resolver ultimately provides a 46 μm linear resolution. A digital to analog converter (DAC) in the computer was programed with a shaped waveform that, after filtering (Ithaco 4302) and amplification (Instruments, Inc. L6), excites a USRD F42 spherical source transducer. The waveform was shaped to provide a maximally flat frequency response over a 3 to 80 kHz band [Fig. 3(b)].

The hydrophone was connected to an Ithaco 1201 preamplifier that provided gains of 20 to 200 times (26 to 46 dB of gain). The signal was bandpass filtered using an Ithaco 4302 filter (250 Hz high-pass and 80 kHz low-pass) and then digitized with a 12-bit analog to digital converter (ADC)

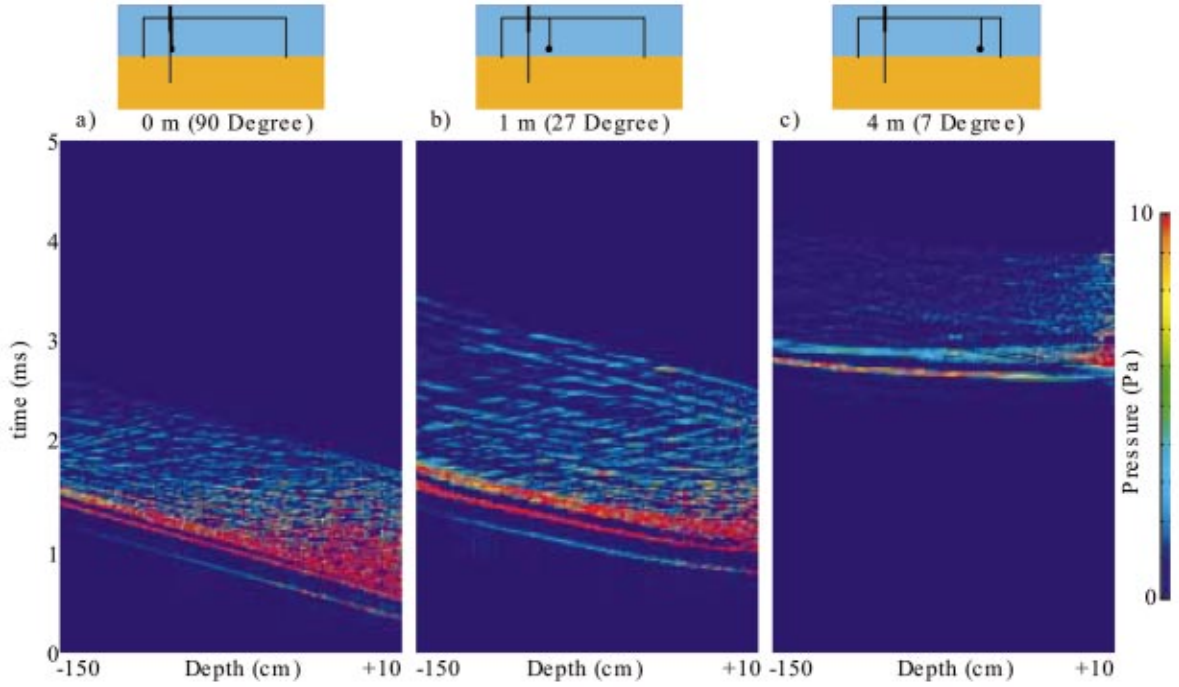


FIG. 4. The averaged temporal response of the synthetic array for three source locations, 0, 1, and 4 m measured horizontally from the insertion point. For all the measurements, the source is located 50 cm above the sand–water interface and the hydrophone is pulled up through the sandy bottom at 1-cm increments. The hydrophone is pulled into the water column an additional 10 cm to allow study of the transition region across the sand–water interface.

residing in the PC. For the measurements reported here, 100 to 200 ensembles were typically recorded for each hydrophone position in the synthetic array.

III. MEASUREMENTS

To calibrate the system after installation, a 1 m free-field reference between the hydrophone and the source was made. The associated temporal and frequency response for a single pulse acquisition is shown in Fig. 3. Also shown [Fig. 3(b)] is the noise floor for this measurement. Figure 3(c) shows the frequency response and noise floor with the hydrophone buried 15 cm and with the source 50 cm directly above the insertion point.

A series of buried hydrophone scans was carried out for different horizontal source positions. Figure 4 shows the averaged temporal responses for the buried synthetic array acquired for three horizontal source locations; (0, 1, and 4 m). These source locations were chosen in relation to the critical angle. The 0 m measurement is well above the critical angle, the 1 m measurement is at the critical angle, and the 4 m measurement is well below the critical angle, measured at the insertion point. Figure 4(a) is the 0 m, normal incidence or 90° measurement, where the insonification angle is measured upward from the sand–water interface. The depth of the sensor (cm) and time (ms) are indicated on the horizontal and vertical axes, respectively. The reverberations from the mounting structure arrive after 1 to 2 milliseconds and have been windowed out of the data set.

The highest energy levels in Fig. 4 are clearly associated with the direct wavefront from the source. However, there are several temporal responses in the data with slopes opposite that of the initial direct wavefront. These are assumed to be associated with scattering from inhomogeneities in the bottom, most likely objects or layers. When the hydrophone was first buried, two or three depths required excessive force to push the water jet through. These depths are presumed to be associated with shell or rock layers.

Also note that there is a dramatic difference between the temporal responses shown in Fig. 4(a) as compared to Fig. 4(c). In Fig. 4(a), the temporal response is linear with hydrophone depth, as would be expected for the source directly above the insertion point. The wave is seen to decrease in amplitude with increasing depth. This is due to two factors, spherical spreading of the wavefront, and attenuation in the sandy bottom. As will be discussed later, the attenuation has a frequency dependence and thus the higher frequencies are attenuated first. As the source is moved from directly over the insertion point to 4 m away, the temporal response of the pressure changes greatly. Figure 4(c) shows an initial single curved line of energy, followed by a second weaker line of energy. Close inspection of this temporal response shows that this earliest arrival [Fig. 4(c)] actually has a very small amplitude near the interface and does not exist above the interface. The temporal pressure responses (Fig. 4) show the pressure from 150 cm below the interface to 10 cm above the interface. The earliest time response in Fig. 4(c) has very little pressure amplitude just under the interface, increases to a maximum amplitude 30 to 50 cm below the interface, and then decreases with depth below 50 cm. Later time arrivals are seen in Fig. 4(c), being strongest just above the interface. This is due to the direct wave in the water column insonifying the array above the sand–water interface. The energy seen in the sandy bottom after the direct wavefront travels to the insertion point [Fig. 4(c)] is a result of a combination of

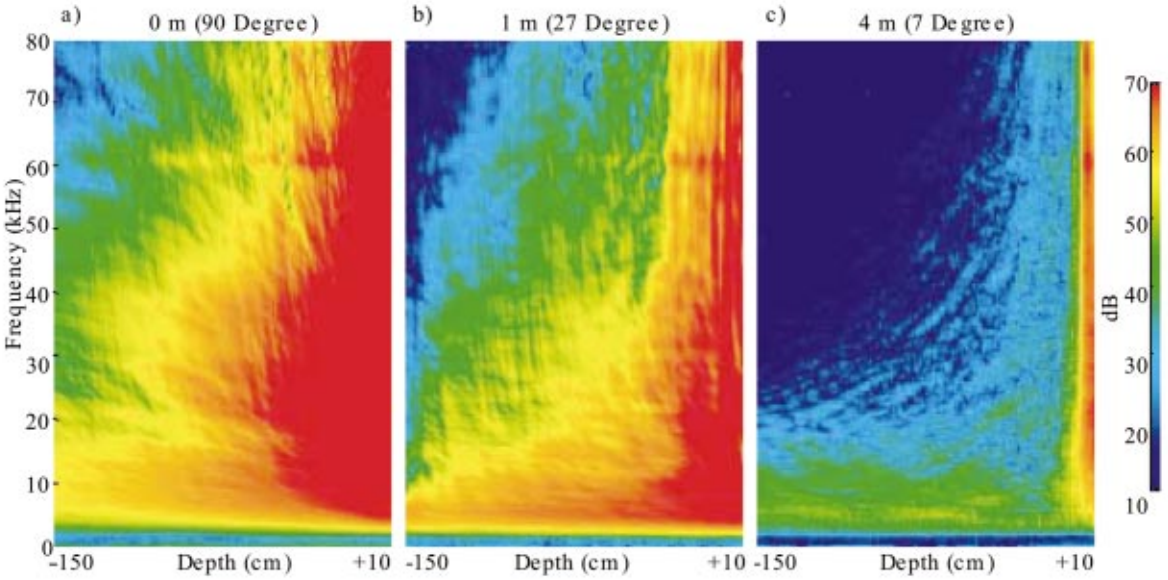


FIG. 5. The frequency response of the synthetic array for three source locations, 0, 1, and 4 m measured horizontally from the insertion point. The vertical axes are frequency in kilohertz and the horizontal axes indicate the depth of the hydrophone from 150 cm below the interface to 10 cm above the interface.

the evanescent field, scattering from surface or volume inhomogeneities, or possibly direct refraction at the sand–water interface of a slower compressional wave. These possibilities will be discussed in detail after further analysis of this data set.

The temporal responses, shown in Fig. 4, were Fourier transformed and the frequency responses of the measurements are shown in Fig. 5. Figure 3 shows the bandwidth of the measurement to be relatively flat from 3 to 80 kHz, even with the hydrophone buried 15 cm below the interface. However, as shown in Fig. 5(a), the pressure decreases with both depth and frequency. This effect is enhanced as the insonification angle, relative to the insertion point, is decreased. Figure 5(b) shows that the pressure at higher frequencies and deeper depths, above 60 kHz and below 100 cm, is below the measurement noise floor. With the source at 4 m (7° insonification angle at the insertion point), the pressure penetrates only a few centimeters into the sediment for frequencies above 20 kHz. However, below 20 kHz, the pressure is still measurable at a depth of 150 cm. The temporal response in Fig. 4(c) and a time-of-flight calculation show that this energy is associated with refraction of the pressure into the sediment near the source and propagation through the sediment to the buried hydrophone. The energy above 20 kHz is attenuated below the measurement noise floor due to the propagation loss through the sand.

In addition to the *in situ* acoustical measurements of this sandy bottom, an analysis of a core sample taken at the measurement site was performed. The core sample consisted of the first 23 cm of the bottom and contained no shell layers. The averaged measured porosity is 0.379. The averaged wet bulk density is 2035 kg/m^3. A sieve analysis shows that 94% to 97% of the aggregate is sand with a mean sieve size of 2.25 ϕ (220 microns).

IV. DISCUSSION

Some advantages of the synthetic array technique employed here have already been illustrated. The power of a densely packed, evenly spaced vertical array can be seen in the postprocessing techniques and analysis. The linearity of the array in the vertical direction supports a direct calculation of the attenuation as a function of frequency from the 0 m (90°) insonification measurement. The phase velocity is also easily established from this same data set. A broader understanding of the pressure in the sand can be obtained from wavenumber-frequency $(k-\omega)$ analysis.

The pressure frequency response as a function of hydrophone depth for the case of the source directly above the insertion point is shown in Fig. 5(a). The phase of the pressure for each frequency bin is taken for each array position and a linear least-squared fit through the data is used to determine the phase speed. Figure 6(a) shows the phase speed as a function of frequency. A slight dispersion in the phase speed can be observed; the phase speed is measured to be ~1710 m/s below 20 kHz and ~1730 m/s above 20 kHz.

The attenuation as a function of frequency is also determined from the 0 m or ''end fire'' measurement, since the spherical wavefront is normal to the array and the calculation of the attenuation can be made directly. Spherical spreading is considered, as described by Urick,[14] and is removed in the attenuation shown in Fig. 6(b). This attenuation also agrees with the data published by Chotiros[15] and calculations given by Turgut and Yamamoto[16] or Ogushwitz.[17] As with the prediction of Turgut and Yamamoto[16] and Ogushwitz,[17] there is a local maximum in the attenuation around 30 kHz.

Turgut and Yamamoto[16] discuss a similar local maximum in the intrinsic attenuation, and use the Kramers–Kronig[18] relationship to check the consistency of the measured compressional wave velocity with the attenuation. Turgut and Yamamoto[16] use the intrinsic attenuation Q^{-1} in their discussion and calculation. The intrinsic attenuation is easily calculated for the measured data taken here. Figure 7(a) shows a comparison between the measured velocity and a calculation based on the Kramers–Kronig relationship. The intrinsic attenuation used for this is shown in Fig. 7(b). The

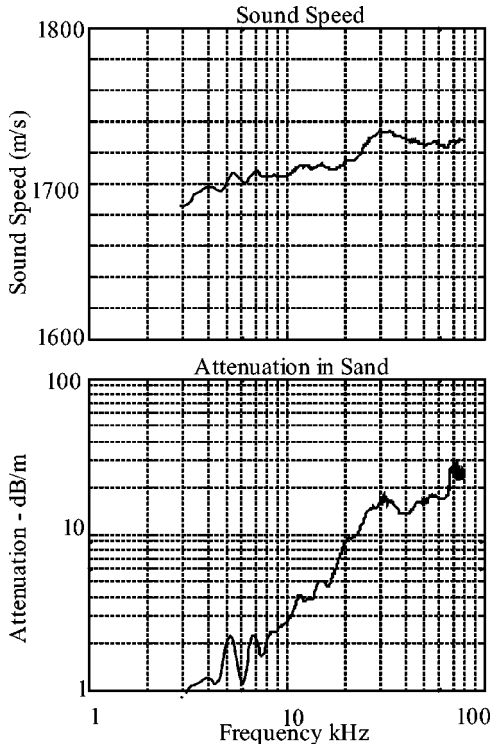


FIG. 6. The averaged sound speed and attenuation in the water-saturated sand. These calculations were made using data taken with the source directly above the insertion point. (a) The sound speed measured using the buried synthetic array. The sound speed is obtained by measuring the phase at each frequency for each hydrophone location. A least-squared linear fit to the phase was used to quantify the averaged sound speed at each frequency. (b) The attenuation calculated from the same measurement. The attenuation is corrected for the spherical spreading from the spherical source. Again, a least-squared fit to the log of the amplitude for each hydrophone location is used to calculate the attenuation.

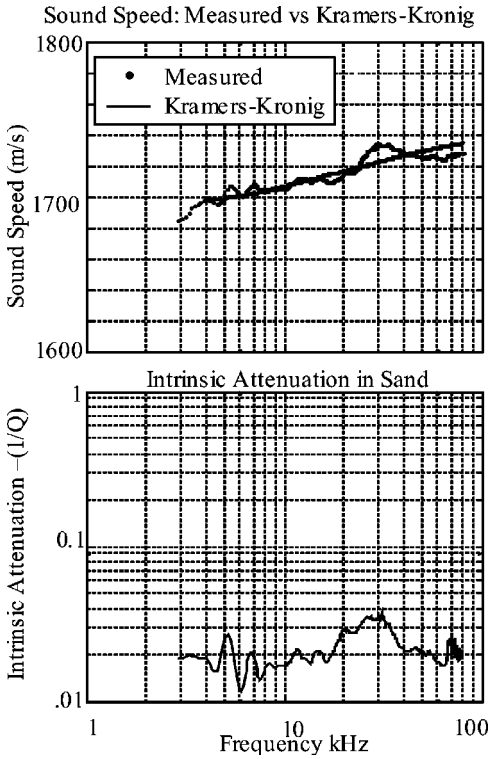


FIG. 7. The sound speed, intrinsic attenuation, and Kramers–Kronig relationship. (a) A comparison between the measured sound speed in the sand and the calculated sound speed based on the intrinsic attenuation and Kramers–Kronig relationship. (b) The intrinsic attenuation (1/Q) measured in the sand.

calculated velocity agrees with the measured velocity and further validates the measurements. Turgut and Yamamoto[16] present data that demonstrate more dispersion than is observed in the data reported here. This larger dispersion results in a more pronounced maximum in the intrinsic attenuation, as related by Kramers–Kronig, than seen in these measurements shown in Fig. 7(b).

The attenuation curve for this data set, shown in Fig. 6(b) and Fig. 7(b), is not a monotonic or a smooth curve as compared to previous pool measurements made by Simpson.[2] For the pool measurements the sand was very uniform, 240-mm mean diameter, and contained no inhomogeneities within the bottom. However, for the measurements made in St. Andrews Bay, the bottoms were sandy, but also included biologics, shells, shell layers, muds, and silts, and could only be considered an inhomogeneous bottom. For these reasons it is expected that the measured attenuation would be a combination of the attenuation associated with an acoustic wave moving through a homogeneous water-saturated sandy bottom and additional attenuation mechanisms such as scattering from the various inhomogeneities within this bottom.

Figure 8 shows the results of wavenumber processing for the three previously studied temporal and frequency data sets displayed in Figs. 4 and 5, respectively. Overlaid on the $k-\omega$ plots shown in Fig. 8 are lines indicating different velocities. The first overlay to be discussed is the dashed white line seen clearly in Figs. 8(b) and (c).

These dashed white lines indicate a velocity associated with a wave moving at 1700 m/s relative to the array. However, there are two variables associated with a wavefront propagating in the sandy bottom, direction and wavespeed. With some understanding of the measurement configuration, additional insight into the wavefront propagating along the array can be obtained. For Fig. 8(a), the source was mounted over the insertion point. It is expected that the majority of the energy measured by the vertical array would propagate in the vertical direction of the array, as indicated by the vertical white and yellow lines that designate possible ray paths of the acoustic wave in the diagram above the data. As is evident in Fig. 8(a), the vast majority of the energy is propagating along the 1700-m/s line and this measurement identifies a specific wave traveling in the sandy bottom. However, additional energy is seen for velocities higher than 1700 m/s, but little can be observed below 1700 m/s. There are two possible explanations for this. One possibility is the presence of wavefronts moving faster than 1700 m/s in this sediment. The other is that energy is impinging on the array at angles different than the vertical direction. This energy could be the result of scattering from inhomogeneities in the sediment or on the sand–water interface. These components are some one to two orders of magnitude below the direct component. For this latter case, the trace velocities for waves would range from 1700 m/s for a wave traveling along the direction of the array, to infinity for a wave traveling normal

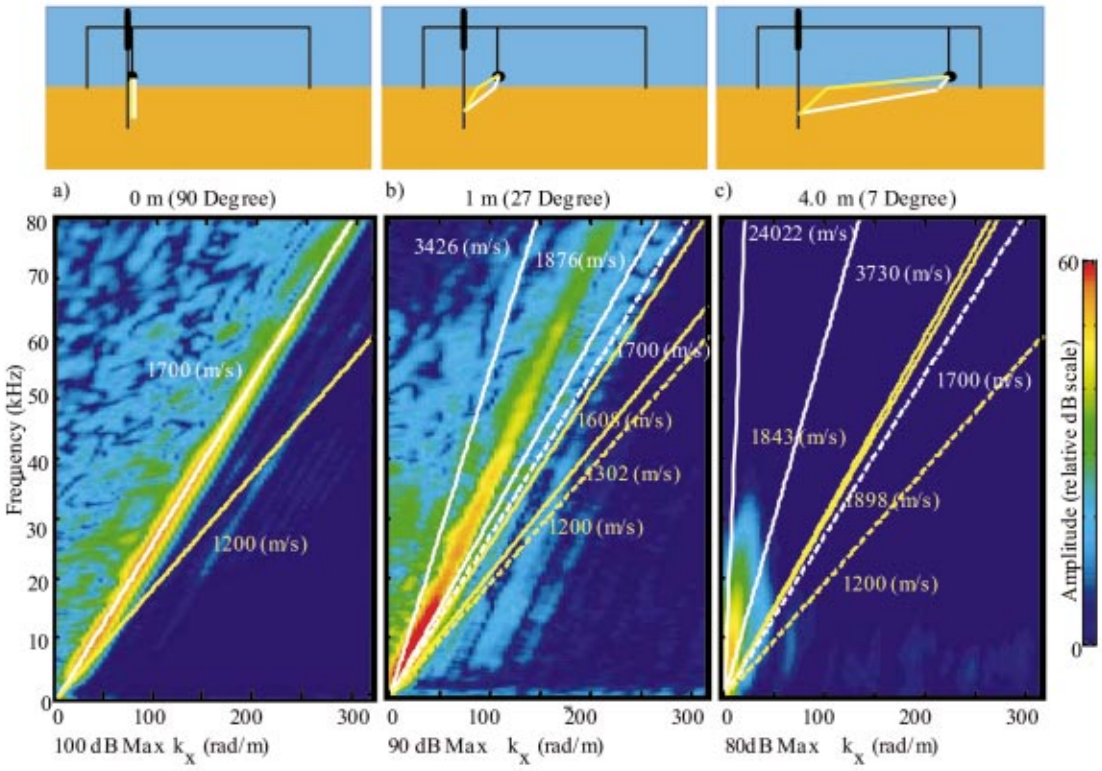


FIG. 8. The wavenumber-frequency response of the synthetic array for three source locations, 0, 1, and 4 m measured horizontally from the insertion point. In the diagrams above the data, the yellow lines indicate possible ray paths for a slow compressional wave entering the sandy bottom, and the white lines indicate possible ray paths for the fast compressional wave entering the bottom. In the wavenumber-frequency space plots, the white and yellow lines bracket the expected trace velocities for a refracted spherical wave with a speed of either 1700 or 1200 m/s, respectively. The white and yellow dashed lines indicate the minimum trace velocities of a fast and slow compressional wave in the sandy bottom, respectively.

to the direction of the array and thus appear with zero wavenumber.

Figures 8(b) and (c) show six lines on them. If a single wave speed in the sand is assumed, for this case 1700 m/s, then the wavenumber plot is easily mapped into the beam pattern of the received energy. Note that for a wave incident on the array at θ measured from the direction of the array relative to the direction of propagation of the wavefront, $\cos\theta = v_s/v_t$, where v_s and v_t are 1700 m/s and the trace velocity, respectively. The white lines are associated with a presumed 1700-m/s wave and the dashed white line indicates the velocity for a 1700-m/s wave propagating vertically along the array. Thus, the direction of the energy impinging on the array is measured directly from the $k-\omega$ plot. Only the positive halves of the wavenumber space are shown in these plots. The diagram above Figs. 8(b) and (c) indicates a single possible ray path of the 1700-m/s wave with white lines and a single possible ray path of a slower 1200-m/s wave with yellow lines.

The upper and lower bounds for the trace velocities, associated with all possible ray paths, can be obtained by modeling the spherical source as a point and by using ray theory to treat the refraction at the interface. In Fig. 8(b) the solid white line overlays indicate the bounds for the trace velocities expected for the second geometry where the source is 1 m horizontally from the insertion point and 50 cm above the interface. Here, it can be seen that the bulk of the energy measured by the array falls in between the two solid white lines. This same observation can be made for the third geometry measured [Fig. 8(c)], where the source is located 4 m horizontally from the insertion point. However, there is lower level energy in all three figures not associated with the direct wavefront.

The manifestation of a slow compressional wave, proposed by Chotiros,[5] and the manifestation of scattering from a roughened interface, predicted by Moe,[8] Thorsos,[9] and Gragg,[10] in the measured data will be considered next. With regard to the former, a second, slower compressional wave in the sandy bottom is assumed to have a wave speed of 1200 m/s, predicted by Chotiros.[5] This 1200-m/s wave speed is indicated by yellow dashed lines in Fig. 8. However, the predicted slower wave would also have a refractive ray path into the sandy bottom for all insonification angles. Thus, for Figs. 8(b) and (c), the associated bounded trace velocities are indicated by the solid yellow lines, where a 1200-m/s wave would appear. Also, for the normal insonification case shown in Fig. 8(a), a 1200-m/s wave would appear as energy along the single yellow line. For insonification angles much less than the critical angle, a 1200-m/s compressional wave would appear between the two solid yellow lines of Fig. 8(c). This trace wave speed for a 1200-m/s wave speed is above the 1700-m/s wave speed for a vertically propagating wave and thus can be confused with a trace velocity of a 1700-m/s wave that is insonifying the array at an angle other than vertical. This surface scattering possibility will be discussed later.

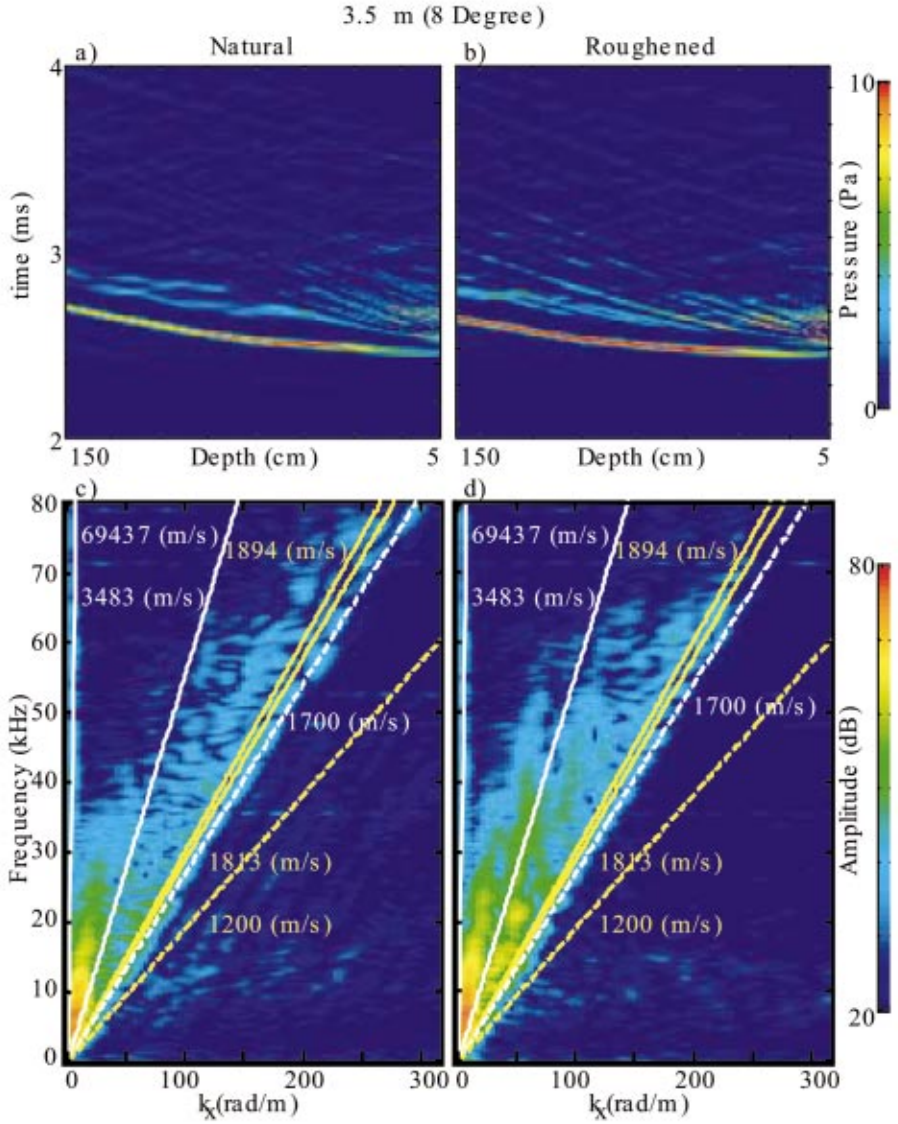


FIG. 9. The temporal and wavenumber-frequency response of the buried synthetic array, with the source 3.5 m measured horizontally from the insertion point, for two interface roughness conditions. (a), (c) The temporal and wavenumber responses for the natural interface condition. (b), (d) The temporal and wavenumber responses for a severely roughened interface condition. Once again, the white and yellow lines bracket the expected trace velocities for a refracted spherical wave with a speed of either 1700 or 1200 m/s, respectively.

In all the plots of Fig. 8 some additional energy is observed below the 1700-m/s line. Note that for energy associated with any wave propagating in the sandy bottom, the wave will have a zero wavenumber at zero frequency and will have an increasing wavenumber as the frequency increases. For a nondispersive wave, the wave speed will follow a straight line on the $k-\omega$ plots. This is important to discriminate coherent energy associated with a traveling wavefield from other processing anamolies. The first processing anomaly easily observed in Fig. 8(b) and to a lesser extent in Fig. 8(a) is associsted with the sidelobes produced from the Fourier transform of finite arrays. The best example of this sidelobe, for both temporal and spatial Fourier transforms, is the sync function that results from a Fourier transform of a rectangular or boxcar window. These sidelobes from the spatial processing of a finite array manifest themselves in the $k-\omega$ data as parallel lines associated with the wave measured by the array. Note in Fig. 8(b) that these sidelobes are very prominent, and parallel to the primary energy measured by the array with a trace velocity of approximately 2400 m/s. These sidelobes, however, do not go through the origin and are thus easily distinguished from the actual trace velocity associated with the traveling wave measured by the array.

Figure 8(a) also shows these sidelobes; however, they are parallel to the 1700-m/s line since for this configuration the source is directly above the array and the velocity measured by the array is the wave speed of the sound in the sand. Note that the three plots in Fig. 8 have the same relative dB range of 60 dB; however, they have a different maximum. This was done to show the data with a good contrast for each set, while still being able to show the large dynamic range needed to display the three measurements.

Within this dynamic range and in all three geometries, no discernible energy is observed for trajectories associated with a trace wave speed below 1700 m/s. Rather, the dominant energy in these displays is due to direct insonification by a refracted 1700-m/s component as well as possible contributions from scattering of the fast compressional wave (1700 m/s) from inhomogeneities or surface roughness.

To better understand specifically how energy scattered from a roughened interface would manifest itself on this syn-

thetic vertical array, a series of measurements was made where the interface was modified. Although there were no quantitative measurements of the surface roughness, for this was not in the original scope of the measurements, a basic study was conducted to understand how surface scattering manifests itself in the data. In addition to a 3.5-m insonification with the natural bottom, which had some intrinsic surface roughness, a second 3.5-m source distance measurement was made with a substantially roughened bottom. A garden hoe was used to put many lines in the sand, several centimeters deep and perpendicular to the source–receiver plane. This greatly enhanced the surface roughness relative to the natural bottom.

Figures 9(a) and (b) show the synthetic array temporal responses for the naturally occurring and roughened interfaces, respectively. For this surface scattering study, the source is 3.5 m away from the insertion point, resulting in an 8° insonification angle measured at the insertion point (well below the critical angle). In both of these displays, the trajectory due to the earliest arrival is clear and distinct. Behind this first arrival is clear evidence of lower level energy arriving between 0.1 ms and 0.5 ms later. Comparison of Figs. 9(a) and (b) shows an increase in the levels of these later arrivals for the artificially roughened surface. As these results are presently understood, the surface roughness causes an increase in the refracted components that appear at later times producing a so-called virtual slow wave.

The $k-\omega$ displays for both cases are shown in Figs. 9(c) and (d), respectively. The amplitude range used in Figs. 9(c) and (d) is the same as that shown in Fig. 8(c), an 80-dB maximum with 60 dB of dynamic range. As with Fig. 8, the trace wave speeds associated with the refraction of 1700-m/s and 1200-m/s waves are indicated by the solid white and solid yellow lines, respectively. The separation in the lines is again associated with the spherical nature of the insonifying wavefronts. Energy is seen between the two solid yellow lines in both Figs. 9(c) and (d). However, the energy levels in Fig. 9(d) show an increase over the energy levels in Fig. 9(c) by about 10–15 dB. Recall that on the $k-\omega$ plots, the direction of the insonifying wavefront is given by the location of the energy on the plot. Thus, most of the increased energy level in Fig. 9(d) falls between the refracted energy of the direct wave and any scattered energy directed down the array. This would be consistent with the interface modifications that were done for this roughened surface scattering measurement.

V. CONCLUSIONS

The concept of synthesizing a vertical array in an at-sea sandy bottom has been demonstrated. We have presented high-quality measurements of sound speed and attenuation over a broad frequency range (3 to 80 kHz). However, since many lengths and resolutions of the synthetic array are possible, this technique can be adapted for lower or higher frequency ranges. For the lower frequency range, below 3 kHz, the hydrophone would have to be buried deeper to increase the aperture of the array. The resolution can be increased by decreasing the sampling distance. For the measurements reported here, the sampling distance was 1 cm, which is sufficient to resolve a 1700-m/s wave up to 80 kHz. The real advantage of this synthetic array is that 200 measurement points over a 2-m vertical line were easily sampled. With a conventional array this same resolution and aperture would require 200 separate hydrophones to be placed in the bottom.

The accuracy of the relative spacing of the array points allows simple processing of the data and provides direct measures of the physical quantity of wave speed and attenuation. Also, since the wavenumber space is easily mapped, the direction of wavefronts can be studied. From the 0 m and 1 m measurement, there is no evidence of a slower wave propagating in this sandy bottom. For the measurement positions greater than 1 m, the wavenumber analysis demonstrates the difficulty in distinguishing roughened interface scattering and the possible existence of a slower 1200-m/s wave. However, these measurements have clearly demonstrated that energy is scattered into the bottom because of interface roughness.

From the temporal data, it is clear that there is acoustic scattering from inhomogeneities in the bottom, possibly shell or rock layers. The broad frequency range of the source provides ample temporal resolution to follow the wavefronts generated due to the scattering from many different sites on the interface and in the bottom. The use of the single hydrophone buried in the bottom minimally disturbs the bottom and the interface, and has no associated self-noise due to reverberations, as exist with multiple hydrophone arrays.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The authors would also like to thank Bill Sawyer for the analysis on the core samples taken at the measurement site.

[1] A. Maguer, E. Bovio, W. L. J. Fox, and H. Schmidt, "*In situ* estimation of sediment sound speed and critical angle," J. Acoust. Soc. Am. **108**, 987–996 (2000).
[2] H. J. Simpson and B. H. Houston, "Synthetic array measurements of acoustical waves propagating into a water-saturated sandy bottom for a smoothed and a roughened interface," J. Acoust. Soc. Am. **107**, 2329–2337 (2000).
[3] A. Maguer, W. L. J. Fox, H. Schmidt, E. Pouliquen, and E. Bovio, "Mechanisms for subcritical penetration into a sandy bottom: Experimental and modeling results," J. Acoust. Soc. Am. **107**, 1215–1225 (2000).
[4] L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed. (Academic, New York, 1980), pp. 5–15.
[5] N. P. Chotiros, "Biot model of sound propagation in water-saturated sand," J. Acoust. Soc. Am. **97**, 199–214 (1995).
[6] K. D. LePage and H. Schmidt, "Spectral integral representations of volume scattering in sediments in layered waveguides," J. Acoust. Soc. Am. **108**, 1557–1567 (2000).
[7] B. H. Tracey and H. Schmidt, "A self-consistent theory for seabed volume scattering," J. Acoust. Soc. Am. **106**, 2524–2534 (2000).
[8] J. E. Moe and D. R. Jackson, "Near-field scattering through and from a two-dimensional fluid–fluid rough interface," J. Acoust. Soc. Am. **103**, 275–287 (1998).
[9] E. I. Thorsos, "The accuracy of perturbation theory for acoustic penetration of sediment due to interface roughness," J. Acoust. Soc. Am. **99**, 2475–2500 (1996).
[10] R. F. Gragg, D. Wurmser, and R. C. Gauss, "Small-slope scattering from rough elastic ocean floors: General theory and computational algorithm," J. Acoust. Soc. Am. **110**, 2878–2901 (2001).
[11] C. R. Bradley and R. S. Stephen, "Modeling of seafloor wave propagation and acoustic scattering in 3D heterogeneous media," J. Acoust. Soc. Am. **100**, 225–236 (1996).

[12] D. D. Ellis and D. M. F. Chapman, ''A simple shallow water propagation model including shear wave effects,'' J. Acoust. Soc. Am. **78**, 2087–2095 (1985).
[13] D. L. Lavoie and M. D. Richardson, ''*In situ* geoacoustic systems and measurements in a variety of carbonate and siliciclastic environments,'' IEEE J. Ocean. Eng. 1467–1473 (1996).
[14] R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1983).
[15] N. P. Chotiros, ''Biot model of sound propagation in water-saturated sand,'' Fig. 13, J. Acoust. Soc. Am. **97**, 199–206 (1995).
[16] A. Turgut and T. Yamamoto, ''Measurements of acoustic wave velocities and attenuation in marine sediments,'' J. Acoust. Soc. Am. **87**, 2376–2382 (1990).
[17] P. R. Ogushwitz, ''Applicability of the Biot theory. III. Wave speeds versus depth in marine sediments,'' J. Acoust. Soc. Am. **77**, 453–464 (1985).
[18] P. M. O'Donnell, E. T. Jaynes, and J. G. Miller, ''Kramers–Kronig relationship between ultrasonic attenuation and phase velocity,'' J. Acoust. Soc. Am. **69**, 696–701 (1981).

The acoustical Klein–Gordon equation: The wave-mechanical step and barrier potential functions[a)]

Barbara J. Forbes[b)]
Phonologica, P.O. Box 43925, London NW2 1DJ, United Kingdom

E. Roy Pike
Department of Physics, King's College London, Strand, London WC2R 2LS, United Kingdom

David B. Sharp
Department of Environmental and Mechanical Engineering, The Open University, Walton Hall, Milton Keynes MK7 6AA, United Kingdom

(Received 18 September 2003; revised 15 May 2003; accepted 19 May 2003)

The transformed form of the Webster equation is investigated. Usually described as analogous to the Schrödinger equation of quantum mechanics, it is noted that the second-order time dependency defines a Klein–Gordon problem. This ''acoustical Klein–Gordon equation'' is analyzed with particular reference to the acoustical properties of wave-mechanical potential functions, $U(x)$, that give rise to geometry-dependent dispersions at rapid variations in tract cross section. Such dispersions are not elucidated by other one-dimensional—cylindrical or conical—duct models. Since Sturm–Liouville analysis is not appropriate for inhomogeneous boundary conditions, the exact solution of the Klein–Gordon equation is achieved through a Green's-function methodology referring to the transfer matrix of an arbitrary string of square potential functions, including a square barrier equivalent to a radiation impedance. The general conclusion of the paper is that, in the absence of precise knowledge of initial conditions on the area function, any given potential function will map to a multiplicity of area functions of identical relative resonance characteristics. Since the potential function maps uniquely to the acoustical output, it is suggested that the one-dimensional wave physics is both most accurately and most compactly described within the Klein–Gordon framework. [DOI: 10.1121/1.1590314]

PACS numbers: 43.20.Ks, 43.20.Mv, 43.70.Bk [MO]

I. INTRODUCTION

According to the comprehensive review of Eisner[1] (1967), the so-called Webster[2] (1919) equation, describing longitudinal wave propagation in a narrow tube, can be traced back to an original derivation by Bernoulli in the 18th century, perhaps to around 1740. This physical system then attracted the attentions of Lagrange, d'Alembert, Euler, and Rayleigh, with closely related problems being addressed by Green and von Helmholtz. It was not until 1946 and the work of Salmon,[3] however, that the simplified form of the equation was finally derived. By taking account of energy conservation, Salmon made a single change of variables and immediately obtained a form mathematically similar to the Schrödinger wave equation of quantum mechanics.

This ''transformed,'' or reduced, form of the equation was most notably treated by Benade and Jansson[4] in 1974. These authors extended the initial physical assumption of plane-wave propagation to one-dimensional wave propagation in general, and they defined ''horn'' functions due to both plane and spherical wavefronts. These expressions were used to give a quantitative account of mode conversion in musical horns.

More recently, Amir[5] (1993) noted that the horn function may be useful in deriving the input impedance of musical horns, and Berners and Smith[6] (1994) and Berners[7] (1999) investigated solutions of the reduced equation in the computational modeling of musical instruments. Investigating unidimensional progagation in flaring horns, Keefe and Barjau[8] (1999) followed the approximate WKB analysis of Benade and Jansson, appropriate to contexts in which the horn function varies slowly over a wavelength.

Recent work by the present authors[9,10] has been examining the reduced Webster equation in further detail, with particular application to the compact parametrization of speech systems.[11,12] Since the reflection and transmission coefficients of the system are frequency dependent, improvements in the convergence characteristics of both the direct and inverse solutions are expected in comparison to standard cylindrical-segment models. Thus, applications in acoustic pulse reflectometry[13] are also being investigated.

The present paper explores some new analytic results in ''wave-mechanical'' acoustics. Section II below briefly recapitulates the reduction of the Webster equation, and finds that the appropriate analogy is with the Klein–Gordon, rather than Schrödinger, equation. It is shown that the duct area function is in a many-to-one mapping with the horn, or ''potential,'' function. Sections III and IV examine the acoustical

[a)]A portion of this work was presented by the first author in ''Piecewise-constant horn functions: The radiation impedance and end corrections in Schrödinger solutions of the Webster equation,'' Proceedings of the International Seminar on Musical Acoustics, Perugia, Italy, pp. 517–520 (Sept. 2001).

[b)]Electronic mail: forbes@phonologica.com

properties of the positive potential functions, the infinite step, and finite barrier. The analysis begins by consideration of the dispersive characteristics of isolated potentials and goes on to quantify their role in duct resonance, with the vocal tract as the paradigmatic case. The wave-mechanical Green's function and a barrier equivalent to a radiation impedance are developed in Sec. IV.

II. THE ACOUSTICAL KLEIN–GORDON EQUATION

The fundamental physical principles of mass and momentum conservation are expressed in terms of an excess pressure, $p'(x,t)$, by the Webster equation

$$\frac{\partial^2 p'(x,t)}{\partial t^2} = c^2 \left\{ \frac{1}{S(x)} \frac{\partial}{\partial x} \left[S(x) \frac{\partial p'(x,t)}{\partial x} \right] \right\}, \tag{1}$$

which describes plane-wave propagation in a duct of slowly varying cross-sectional area, $S(x)$, where c is the speed of sound in the ambient medium (neglecting dissipative losses). Comparison with exact solutions[14] gives an indication of its validity. Equation (1) has been derived many times in the literature on the basis of classical and linearizing assumptions; briefly, these are that the fluid is compressible and ideal (homogeneous and nonviscous), disturbances are adiabatic and of low Mach number, p' is small compared to the bulk modulus of the medium, and that the tract walls are rigid. It is known that corrections to the assumptions of nonviscous and adiabatic propagation[15] are required for tracts of cross section less than ~ 0.2 cm^2, but the role of the duct geometry in this transition—and to, for example, a high Reynolds number and turbulent flow—remains unquantified (although we have suggested[9] that it may be marked by the introduction of bound-state solutions of the reduced Webster equation).

Conversely, if circular symmetry is assumed and evanescent modes[14] are neglected, an upper bound on the cross section can be set in terms of the cut-on frequency, f_c, of the first higher mode, defined in terms of the radius, r, as

$$f_c = \frac{3.83c}{2\pi r}. \tag{2}$$

In speech production, for example, for a maximal vocal-tract cross section of around 15 cm^2 and the sound velocity 350 ms^{-1}, plane-wave propagation will hold up to around 9.8 kHz, and up to 4.7 kHz if asymmetric modes are admitted.[16] This encompasses the main speech range and suggests that speech acoustics is, to a first approximation, largely independent of mode conversion phenomena.

It was Salmon[3] who first noted that the pressure and area functions, $p'(x,t)$ and $S(x)$, in the Webster equation must together obey the principle of conservation of energy such that, averaged over a period, τ

$$\langle p'^2(x,t) \rangle_\tau S(x) = \text{const.} \tag{3}$$

Defining a new variable, the slowly varying "acoustical wave function," $\Psi(x,t)$, as

$$\Psi(x,t) = p'(x,t) S(x)^{1/2}, \tag{4}$$

thus removes much of the predictable fluctuation of pressure with axial distance and elucidates the variations in phase velocity. It can be noted that, in contrast to the quantum-mechanical case, the acoustical wave function is a measurable quantitity; and, since the potential energy per unit volume of fluid is defined as $|p'(x,t)|^2/(2\rho_0 c^2)$, then $|\Psi|^2$ is directly proportional to the potential energy per unit length of fluid.[17]

Substitutions within the Webster equation then result in the reduced form

$$\frac{\partial^2 \Psi(x,t)}{\partial t^2} = c^2 \left\{ \frac{\partial^2 \Psi(x,t)}{\partial x^2} - U(x) \Psi(x,t) \right\}, \tag{5}$$

here referred to as the "Klein–Gordon" form because of its similarity to the relativistic wave equation for a massive spinless particle,[18] and also to the equation for a string embedded in an elastic medium.[19] In particular, the second-order time derivative distinguishes it from the Schrödinger equation, which is first order in this respect. Some interesting implications of the derivative have been noted by Berners,[7] in regard to noncausal errors in the conical discretization of ducts.

Equation (5) has the form of a wave equation appropriate to the assumptions holding for the Webster equation, and the function $U(x)$ is defined in terms of the duct area function, as

$$U(x) = \frac{1}{2S(x)} \left\{ \frac{d^2 S(x)}{dx^2} - \frac{1}{2S(x)} \left[\frac{dS(x)}{dx} \right]^2 \right\}, \tag{6}$$

or

$$U(x) = \frac{d^2 S(x)^{1/2}/dx^2}{S(x)^{1/2}}. \tag{7}$$

Previous authors have noted that $U(x)$ plays a similar role to the potential energy function of quantum mechanics, and that complete information about the shape of the tract, and hence the reflection and transmission characteristics of the system, is contained therein. Thus, it will continue to be referred to throughout as the "potential function," or "potential," in analogy with the wave-mechanical case, although clearly, it defines a geometry rather than an energy. Benade and Jansson extended the domain of the Webster equation from plane-wave propagation to one-dimensional flow in general, noting that the potential function vanished for both straight tubes and cones. They went on to discuss the potential-function characteristics of spherical wavefronts, examining, for instance, duct regions where a conical segment was matched to a flaring bell. In such cases, $S(x)$ in Eq. (7) above should be identified with the area of the spherical cap. For circular symmetry and the special approximation of a plane wave, however, $S(x)^{1/2}$ is directly proportional to the tract radius, $r(x)$, and the potential function can be described as a "scaled curvature" of the duct wall. For convenience, this interpretation will be made in the present work.

Since $U(x)$ does not depend on time, a solution

$$\Psi(x,t) = \psi(x)\phi(t), \tag{8}$$

is assumed. By separation of variables it is found that

$$\frac{d^2\phi(t)}{dt^2} = -\omega^2\phi(t), \tag{9}$$

and

$$\frac{d^2\psi(x)}{dx^2} + [k^2 - U(x)]\psi(x) = 0, \tag{10}$$

where k is the free-space wave number, neglecting dissipation, and $\omega = ck$.

The second-order time dependency yields, in contrast to the Schrödinger equation, two solutions of Eq. (9), of the form

$$\phi(t) = Ae^{-i\omega t} + Be^{i\omega t}, \tag{11}$$

in terms of generally complex coefficients. For $U(x)$ piecewise constant, such that $U(x) = U_0$, the solutions[18] of Eq. (10)

$$\psi(x) = Ce^{-i\hat{k}x} + De^{i\hat{k}x}, \tag{12}$$

are in terms of a dispersive wave number $\hat{k} = (k^2 - U_0)^{1/2}$. Evanescent, or exponentially decaying, solutions exist for the case $U_0 > k^2$. Otherwise, the dispersion is such that the wave number is reduced by a positive value of U_0, and increased by a negative value, resulting in increased and decreased phase velocities, $c' = \omega/\hat{k}$, respectively. Dispersive solutions of this type are unique to the acoustical Klein–Gordon equation: one-dimensional cylindrical and conical duct models[20] assume dispersionless propagation.

It may now be noted that a unique underlying area function can be found from any particular potential function by solution of Eq. (7) as a second-order homogeneous differential equation

$$\frac{d^2S(x)^{1/2}}{dx^2} - U(x)S(x)^{1/2} = 0, \tag{13}$$

only by obtaining a particular solution in terms of two known initial conditions on the area function. For piecewise-constant, or ''square,'' potentials, a general solution of Eq. (13) can be obtained in terms of the two special cases (i) $U_0 > 0$ and (ii) $U_0 < 0$. The former case of a ''positive'' potential function is appropriate to the wave-mechanical step and barrier; the latter ''negative'' potential function describes the wave-mechanical well.

A. (i) $U_0 > 0$

It is seen from Eq. (7) that the condition $U_0 > 0$ corresponds to a positive curvature, such as is found in the mathematical horns and in the bell of musical horns. Equation (13) can be written as

$$\frac{d^2S(x)^{1/2}}{dx^2} - U_0S(x)^{1/2} = 0, \tag{14}$$

with corresponding roots $\lambda_1 = U_0^{1/2}$, $\lambda_2 = -U_0^{1/2}$ and the general, hyperbolic, solution

$$S(x)^{1/2} = C_1e^{U_0^{1/2}x} + C_2e^{-U_0^{1/2}x}. \tag{15}$$

For an initial-value problem and setting $A_0 = S(0)^{1/2}$ and $\xi = dS(0)^{1/2}/dx$, particular solutions can be obtained in terms of

$$C_1 = \frac{1}{2}\left(\frac{\xi}{U_0^{1/2}} + A_0\right), \tag{16}$$

and

$$C_2 = \frac{1}{2}\left(A_0 - \frac{\xi}{U_0^{1/2}}\right). \tag{17}$$

For zero initial gradient, Eq. (15) reduces to

$$S(x)^{1/2} = A_0 \cosh U_0^{1/2}x, \tag{18}$$

describing a section of catenoidal horn.

B. (ii) $U_0 < 0$

Equation (7) entails that the acoustical well corresponds to a negative curvature, most obviously, to a constriction. Equation (13) now has the form

$$\frac{d^2S(x)^{1/2}}{dx^2} + |U_0|S(x)^{1/2} = 0, \tag{19}$$

with corresponding roots $\lambda_3 = i|U_0|^{1/2}$, $\lambda_4 = -i|U_0|^{1/2}$, and the general, trigonometric, solution

$$S(x)^{1/2} = C_3 \cos|U_0|^{1/2}x + C_4 \sin|U_0|^{1/2}x. \tag{20}$$

It is found that

$$C_3 = A_0, \tag{21}$$

and

$$C_4 = \frac{\xi}{U_0^{1/2}}. \tag{22}$$

For the case of zero initial gradient, then

$$S(x)^{1/2} = A_0 \cos|U_0|^{1/2}x, \tag{23}$$

describing an area function that vanishes with axial distance and with closure at

$$x = \frac{\pi}{2|U_0|^{1/2}}. \tag{24}$$

Particular solutions for $A_0 = 1.0$ cm and zero initial gradient, $\xi = 0$, are illustrated schematically in Fig. 1 for unit potential-function width and height ($U_0 = 1\ \mathrm{m}^{-2}$), and in Fig. 2 for unit depth ($U_0 = -1\ \mathrm{m}^{-2}$).

A complete, continuous area function may be recovered from just one, or any sequence of, square potentials of arbitrary height and width, just from two initial conditions and the subsequent matching of particular solutions and their first derivatives at the discontinuities of U_0. This process amounts to the linking of hyperbolic and trigonometric regions of positive and negative potential, respectively, by segments of cylindrical or conical tube, for which $U_0 = 0$. In Fig. 1, for example, whereas the region of zero potential to the

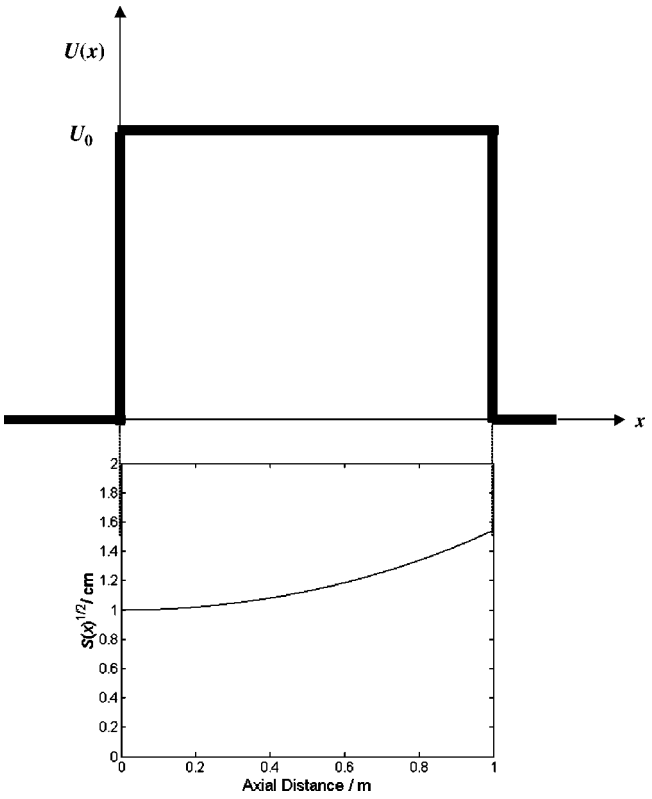


FIG. 1. A barrier potential function of unit width and height, $U_0=1\ \mathrm{m}^{-2}$ (top), and a particular solution of $S(x)^{1/2}$ for $\xi=0$.

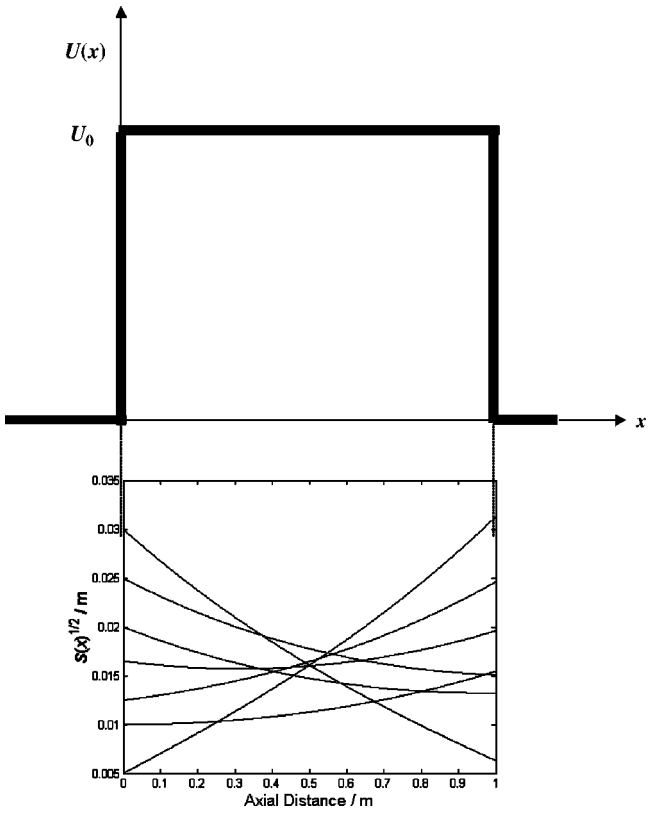


FIG. 3. Arbitrary members of the equivalence set of a barrier potential function of unit width and height, $U_0=1\ \mathrm{m}^{-2}$.

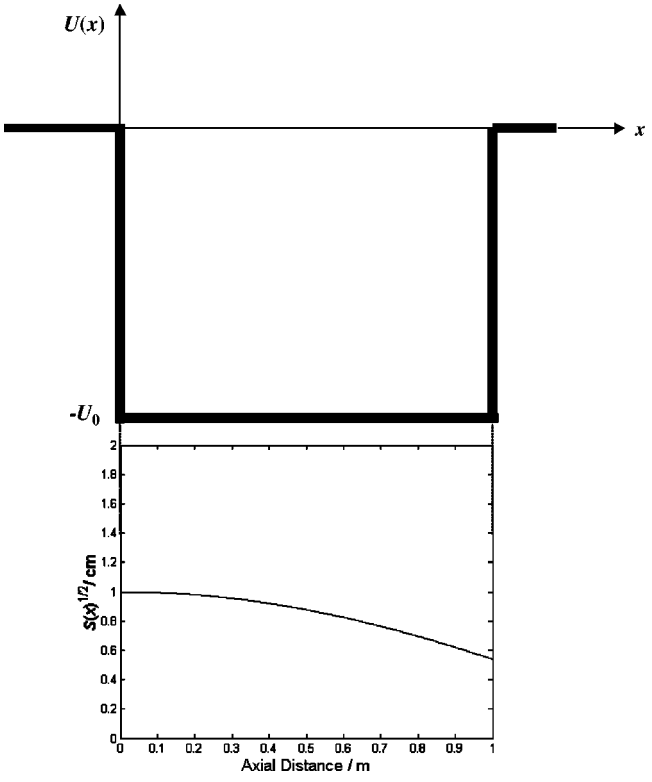


FIG. 2. A well potential function of unit width and depth, $U_0=-1\ \mathrm{m}^{-2}$ (top), and a particular solution of $S(x)^{1/2}$ for $\xi=0$.

left of the barrier corresponds to an initially straight duct, the region extending from the right corresponds to a smoothly matched, infinitely long cone.

For arbitrary initial conditions, any particular potential function will, in fact, map to an infinite ''equivalence set'' of area functions. The dispersive properties of this set are identical, and in Sec. IV, it will be shown that their reflection, transmission, and relative resonance characteristics are also identical. This is the pipe analog of the ''drum'' problem, famously posed by the mathematician, Kac.[21]

Further members in the equivalence sets containing the area functions of Figs. 1 and 2 are shown in Figs. 3 and 4, for arbitrary initial conditions. It is clear that, depending on the value of ξ, a barrier potential may correspond to a region of decreasing radius, that is, to a constriction, despite the positive curvature. Similarly, a well may correspond to a region of overall expansion, although the curvature is negative.

Although the acoustical signal is evidently determined by subtleties in the potential function, $U(x)$, rather than the area function *per se*, the potential-function formalism is not widely referred to in the acoustics literature, and not at all in speech acoustics. Thus, although the basic potential function types are well known from the quantum-mechanical literature, their acoustical analogs are largely unexplored. The present work deals with the positive potentials, the infinite step, and finite barrier. The acoustical characteristics of the wave-mechanical well[9,10,12] will be explored in the second paper of this series.

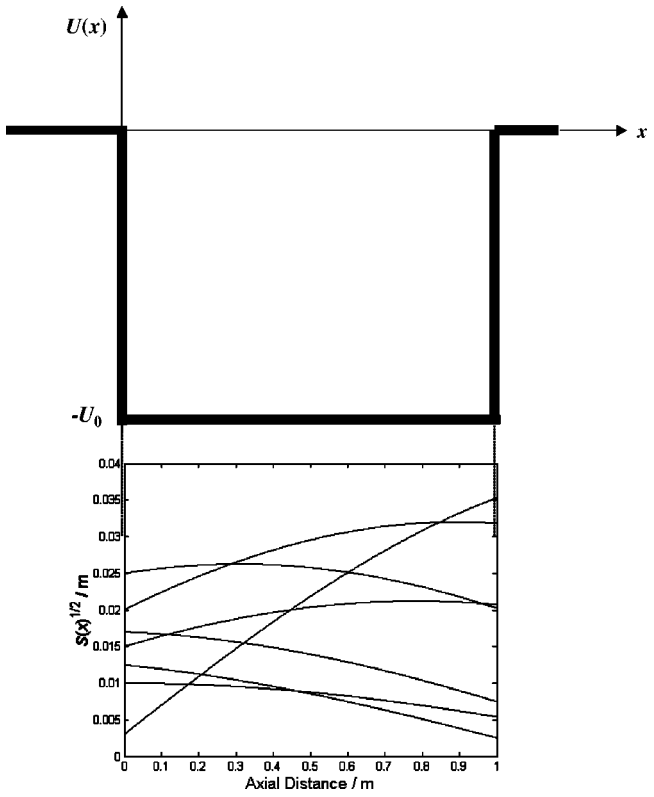


FIG. 4. Arbitrary members of the equivalence set of a well potential function of unit width and depth, $U_0=-1\ \mathrm{m}^{-2}$.

III. CLASSICAL THEORY OF ACOUSTICAL POTENTIALS: THE INFINITE STEP

A. Pipe resonance: A review

The standard method of determination of pipe resonances begins by the consideration of acoustic impedances. For a standing wave in a pipe of constant cross section S, radius r, and length l, the plane-wave[15] acoustic impedance, Z, is defined in terms of the excess incident and reflected pressures, $p'(x,t)$, and volume velocities, $u'(x,t)$, as

$$Z=\frac{p_i'+p_r'}{u_i'+u_r'}, \tag{25}$$

and thus by the Euler equation, the general expression

$$Z=\frac{\rho_0 c}{S}\frac{Ae^{-ikx}+Be^{ikx}}{Ae^{-ikx}-Be^{ikx}}, \tag{26}$$

is obtained for ρ_0, the equilibrium density of the ambient medium.

It is convenient to define the origin, $x=0$, at the termination to free space, and the duct input at $x=-l$. Writing the terminating impedance, Z_0, as a sum of real (resistive) and imaginary (reactive) parts

$$Z_0=\frac{\rho_0 c}{S}(R_0+iX_0), \tag{27}$$

and the input impedance, Z_{-l}, similarly, it is straightforward to show that

$$R_{-l}=\frac{R_0(1+\tan^2 kl)}{(1-X_0\tan kl)^2+R_0^2\tan^2 kl}, \tag{28}$$

and

$$X_{-l}=\frac{X_0-(X_0^2+R_0^2-1)\tan kl-X_0\tan^2 kl}{(1-X_0\tan kl)^2+R_0^2\tan^2 kl}. \tag{29}$$

In the case of an infinite baffle at $x=0$, and for small X_0, $R_0\ll X_0$, and $kr\ll 1$, the terminating impedance can be approximated at the minimum order as

$$Z_0=\frac{\rho_0 c}{S}\left(\frac{k^2r^2}{2}+\frac{i8kr}{3\pi}\right), \tag{30}$$

leading to a frequency-independent end correction, such that the effective length of the pipe is $l_{\mathrm{eff}}=l+8r/3\pi$. Typical vocal-tract parameters of $S=5\ \mathrm{cm}^2$ and $l=17.5\ \mathrm{cm}$, and the assumption of a high-impedance source, yield quarter-wavelength resonances at

$$k=\frac{(2n+1)\pi}{2l_{\mathrm{eff}}},\quad n=0,1,2\ldots. \tag{31}$$

Formants are obtained at 470,1410,2350,... Hz, that is, at around 94% of the nominal values.

In the next section, we examine the effect of a Klein–Gordon potential function on the acoustic impedance and on duct resonance. An analytic expression for a frequency-dependent end correction is derived for the special case of a long, hyperbolic, baffle. It is found to be related to the wave-mechanical penetration of the plane wave into a ''step'' potential. Although not exact with respect to the 3D wave equation, the result is of interest since, until recently,[22] theoretical values have been available only for the extreme cases of an unflanged cylinder, and one terminated in an infinite baffle.

B. Determination of the resonances of a duct terminated in a step potential

Benade and Jansson previously noted that the plane-wave approximation gives inaccurate results for the power transmitted across a finite positive potential, or barrier, since the swelling of the wavefront due to the expansion is unaccounted for. In this section, an account is given of the range of validity of the Klein–Gordon equation by calculation of the plane-wave penetration depth in the radiationless case, beyond which higher-order mode solutions should be considered.

Consider an infinitely long barrier, the ''step potential'' (Fig. 5). Where propagation is below the cut-on frequency at $k=U_0^{1/2}$, the infinite-step configuration corresponds to a tract of zero curvature terminated in a hyperbolic baffle, long such that there is no plane-wave transmission. A particular instance, illustrated schematically in Fig. 6, is that of an organ pipe terminated in an infinite catenoidal horn. The case then has essential similarities to that of the exponential-horn loudspeaker. In this section, the ideal case of zero transmission is discussed; in Sec. IV below, it is shown that a barrier of a just a few millimeters thickness has similar characteristics.

The potential function, $U(x)$, can be described as

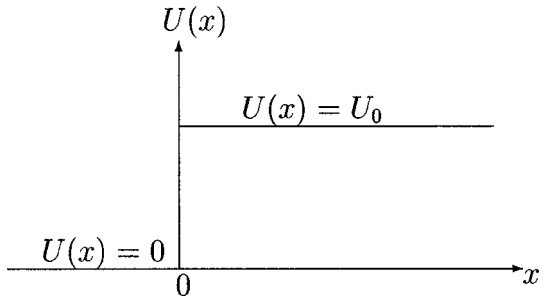


FIG. 5. The step potential.

$$U(x)=\begin{cases} 0 & x<0 \\ U_0 & 0<x, \end{cases} \tag{32}$$

and the following solutions are assumed:

$$\psi(x)=\begin{cases} Ae^{-ikx}+Be^{ikx} & x<0 \\ Ce^{-i\hat{k}x}+De^{i\hat{k}x} & x>0, \end{cases} \tag{33}$$

where $\hat{k}=(k^2-U_0)^{1/2}$. The dispersive wave number, $\hat{k}$, is imaginary in the case of $k^2<U_0$, but real, and reduced, where propagation is above the barrier.

Calculation of the reflection and transmission coefficients presents two cases according to the barrier height in relation to the incident wave number.

1. $k^2<U_0$

Following the discussion in Sec. II, which noted that plane-wave propagation can be assumed up to around 5 Hz for typical vocal-tract cross sections, it is found that the condition $k^2<U_0$ defines a minimum barrier height of around $9\times10^3\,\mathrm{m}^{-2}$. An exponentially increasing solution is obtained as the wave number becomes imaginary for $x>0$ but, since there is no reflected solution for $x>0$ due to the infinite barrier width, we may choose to set D to zero for conservation of energy by defining $\hat{k}\equiv -ik'$, where $k'=(U_0-k^2)^{1/2}$. Equating solutions and their derivatives at $x=0$, the identities

$$A+B=C, \tag{34}$$

and

$$A-B=-\frac{ik'}{k}C, \tag{35}$$

are obtained. Adding and subtracting (34) and (35) gives, respectively

$$A=\frac{C}{2}(1-i\gamma), \tag{36}$$

and

$$B=\frac{C}{2}(1+i\gamma), \tag{37}$$

where

$$\gamma=\frac{k'}{k}. \tag{38}$$

It is known[8] that, below cut-on, the acoustic impedance at the throat of an infinite hyperbolic horn is purely reactive, and that

$$X_0=\frac{k}{k'+\alpha}, \tag{39}$$

where

$$\alpha=\frac{dS(0)/dx}{2S(0)}. \tag{40}$$

For a catenoidal horn, therefore, since $\alpha=0$, it is possible to define an end correction, Δl, as

$$\Delta l=k^{-1}\arctan(\gamma^{-1}), \tag{41}$$

which, at low frequencies, can be approximated as

$$\Delta l=1/k'. \tag{42}$$

Now, turning to the combined system of Fig. 6, substitution of the reflection coefficients, A, B, above into Eq. (26) yields

$$R_{-l}=0, \tag{43}$$

and

$$X_{-l}=\frac{1+\gamma\tan kl}{\gamma-\tan kl}. \tag{44}$$

The input impedance, Z_{-l}, is purely reactive, as would be expected in the absence of radiation. The tract resonances due to a high-impedance source, that is, for a closed–open tube, occur for the poles of X_{-l}, namely for

$$\tan kl=\gamma, \tag{45}$$

and lower from their quarter-wavelength positions with decrease in step height. As U_0 is lowered from $U_0\to\infty$ to $U_0=1.0\times10^5\,\mathrm{m}^{-2}$, for example, a graphical solution of Eq. (45) shows that the poles of X_{-l} for a tract of nominal length

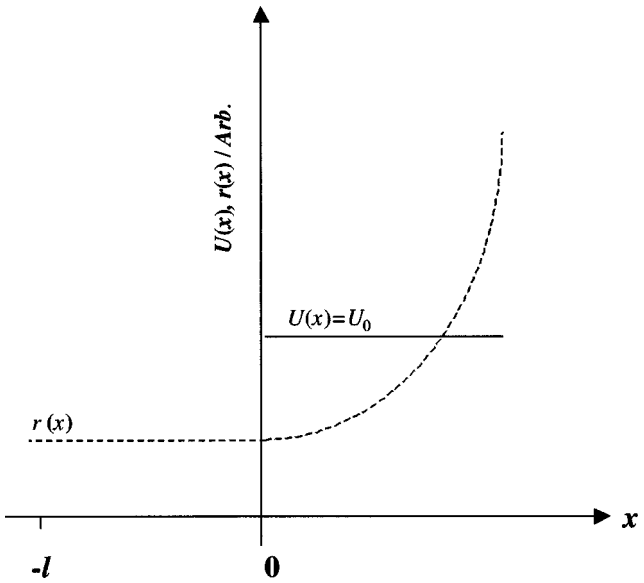


FIG. 6. The step potential and a particular solution for the area function (schematic).

$l=17.5$ cm fall to 491, 1473, 2455, 3437 and 4419±1 Hz.

As $U_0\to\infty$, or for $k^2\ll U_0$, $\gamma\to\infty$ and the determination of the duct resonances reduces to a Sturm–Liouville[7] eigenvalue problem, appropriate to homogeneous boundary conditions.

Equation (44) also admits the half-wavelength eigenvalues of an open–open tube for $\tan kl=-\gamma^{-1}$, for which $X_{-l}=0$.

Solutions may also be approximated analytically on the assumption of small shifts δk from the nominal quarter-wavelength eigenvalues. For $\delta k\ll k$ and working around $k_n=(2n+1)\pi/2l$, corrected tract eigenvalues, or formants, F_n, are obtained as

$$F_n\approx\frac{c}{2\pi}\left\{\frac{-\arctan(\gamma_n^{-1})+n\pi}{l}\right\},\quad n=0,1,2...,\tag{46}$$

where

$$\gamma_n=\left\{\frac{4U_0l^2}{(2n+1)^2\pi^2}-1\right\}^{1/2}.\tag{47}$$

The effective length of the tract is, clearly, a function of the step height, U_0. A frequency-dependent end correction is identified as a function of the Klein–Gordon potential function, the tract curvature scaled by the radius [Eq. (7)], and so is shown to be sensitive to particular aspects of the duct geometry other than the terminating radius *per se*.

Setting U_0 to $1.0\times10^5\,\mathrm{m}^{-2}$ gives cut-on at around four times above the plane-wave limit and causes the formants of the standard tract to drop to approximately 98% of their nominal values, that is, to 491, 1473, 2455, 3436, and 4418 Hz for F_1–F_5, in full agreement with previous results. The appropriate end corrections are 3.21, 3.22, 3.23, 3.24, and 3.26 mm, respectively.

These results are in agreement to better than 1.6% with those obtained from Eq. (42), which yields 3.16 mm at nominal F_1 and 3.27 mm at F_5. Furthermore, Eq. (42) can be identified with the analytic expression for wave function penetration into a step potential, obtained by noting that the amplitude of the evanescent wave, $\sqrt{C^*C}e^{-k'x}$, will fall rapidly for $x>1/k'$. The wave-mechanical penetration depth is therefore equal to $1/k'$ and, in the low-frequency limit, is approximately frequency independent and equal to $U_0^{-1/2}$. For the general case of a hyperbolic baffle, corrections to these values can be made [Eq. (40)] in terms of initial values on the area function.

2. $k^2>U_0$

The reflection and transmission coefficients are now defined as $\hat{k}\equiv k''$, where $k''=(k^2-U_0)^{1/2}$; that is, the wave propagates with a longer wavelength above the barrier. By the previous method, it is found that

$$R_0=\frac{k}{k''},\tag{48}$$

and

$$X_0=0,\tag{49}$$

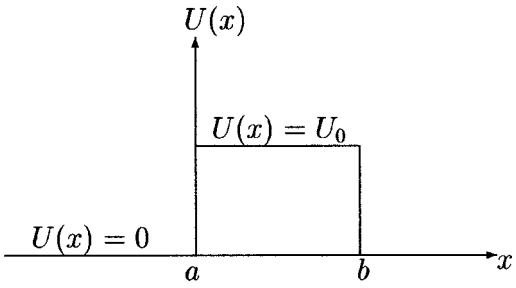


FIG. 7. The barrier potential.

so that the terminating impedance is purely resistive. For $k^2\gg U_0$, $R_0\to1$ and $Z_0\to\rho_0c/S$, as required in the high-frequency limit.

This hypothetical case is included for completeness since it is applicable only to low barriers and cannot be approximated by a barrier of finite width.

Positive potentials of finite width are now discussed, recalling the principal result of this section, namely that substantial wave function transmission will occur for widths less than the plane-wave penetration depth taken at a particular step height. This was found to be approximately 3 mm at $U_0=1.0\times10^5\,\mathrm{m}^{-2}$. Widening a barrier will lead to progressive underestimation of the transmission unless a separate attempt is made to account for higher-order mode phenomena.

IV. CLASSICAL THEORY OF ACOUSTICAL POTENTIALS: THE BARRIER

In the previous section, an analytic expression was obtained for a frequency-dependent end correction in the radiationless case of an acoustical duct terminated in an infinite hyperbolic baffle, the wave-mechanical ''step.'' For cases of substantial energy transmission, such simple expressions are not forthcoming and, as for the boundary conditions previously noted by Bonder,[23] the acoustical context defines a ''resonance,'' rather than eigenvalue, problem. In this section, therefore, an appropriate analytic methodology is developed in terms of the Green's function for the duct impulse response. As noted by Benade and Jansson, the boundary condition at the termination to free space constitutes a singularity in the plane-wave potential function: this section shows, however, that for low-frequency applications such as vowel simulations, it is possible to approximate the pole at the mouth by a barrier of finite dimensions (Fig. 7).

The barrier function, $U(x)$, can be described as

$$U(x)=\begin{cases}0 & x<a\\ U_0 & a<x<b\\ 0 & b<x,\end{cases}\tag{50}$$

and the solutions are in general

$$\psi(x)=\begin{cases}Ae^{-ikx}+Be^{ikx} & x<a\\ Ce^{-i\hat{k}x}+De^{i\hat{k}x} & a<x<b\\ Ee^{-ikx}+Fe^{ikx} & x>b,\end{cases}\tag{51}$$

for a dispersive wave number, $\hat{k}=(k^2-U_0)^{1/2}$, as before. It is important to note that, since the barrier is of finite width,

both forward- and backward-propagating solutions exist within it, in contrast to the case of the infinite step where there was no reflected solution beyond $x=0$. However, by the same reasoning, if there is no barrier beyond $x=b$, F may be set to zero and the wave functions and their first derivatives may be matched at the barrier edges, giving reflection and transmission coefficients by the method of Sec. III B above. It is now more appropriate, in fact, to introduce a transfer-matrix method of calculating the propagation coefficients, that generalizes completely to arbitrary sequences, or ''strings,'' of square potential functions.

A. Transfer-matrix method of calculating reflection and transmission coefficients

It is convenient to put $\hat{k}_i=(k^2-U_{0i})^{1/2}$ for the ith barrier in a sequence $i=0,1,...,n$, of height U_{0i} and width $\Delta_i=b_i-a_i$. Matching wave functions at $x=a_i$, it is found that

$$A_i e^{-ika_i}+B_i e^{ika_i}=C_i e^{-i\hat{k}_i a_i}+D_i e^{i\hat{k}_i a_i}, \tag{52}$$

and taking derivatives

$$-ikA_i e^{-ika_i}+ikB_i e^{ika_i}=-i\hat{k}_i C_i e^{-i\hat{k}_i a_i}+i\hat{k}_i D_i e^{i\hat{k}_i a_i}. \tag{53}$$

Multiplying (52) by ik throughout, and subtracting and adding (53), it is found that

$$A_i=\frac{C_i}{2k}(k+\hat{k}_i)e^{-i(\hat{k}_i-k)a_i}+\frac{D_i}{2k}(k-\hat{k}_i)e^{i(k+\hat{k}_i)a_i}, \tag{54}$$

and

$$B_i=\frac{C_i}{2k}(k-\hat{k}_i)e^{-i(k+\hat{k}_i)a_i}+\frac{D_i}{2k}(k+\hat{k}_i)e^{i(\hat{k}_i-k)a_i}. \tag{55}$$

Similarly, at $x=b_i$ and multiplying by $i\hat{k}$, the following coefficients are obtained:

$$C_i=\frac{E_i}{2\hat{k}_i}(k+\hat{k}_i)e^{-i(k-\hat{k}_i)b_i}+\frac{F_i}{2\hat{k}_i}(\hat{k}_i-k)e^{i(k+\hat{k}_i)b_i}, \tag{56}$$

and

$$D_i=\frac{E_i}{2\hat{k}_i}(\hat{k}_i-k)e^{-i(k+\hat{k}_i)b_i}+\frac{F_i}{2\hat{k}_i}(k+\hat{k}_i)e^{i(k-\hat{k}_i)b_i}. \tag{57}$$

Two matrix identities are obtained for the general case

$$\begin{pmatrix}A_i\\B_i\end{pmatrix}=\frac{1}{2k}\begin{pmatrix}(k+\hat{k}_i)e^{-i(\hat{k}_i-k)a_i} & (k-\hat{k}_i)e^{i(k+\hat{k}_i)a_i}\\(k-\hat{k}_i)e^{-i(k+\hat{k}_i)a_i} & (k+\hat{k}_i)e^{i(\hat{k}_i-k)a_i}\end{pmatrix}\begin{pmatrix}C_i\\D_i\end{pmatrix}, \tag{58}$$

and

$$\begin{pmatrix}C_i\\D_i\end{pmatrix}=\frac{1}{2\hat{k}_i}\begin{pmatrix}(k+\hat{k}_i)e^{-i(k-\hat{k}_i)b_i} & (\hat{k}_i-k)e^{i(k+\hat{k}_i)b_i}\\(\hat{k}_i-k)e^{-i(k+\hat{k}_i)b_i} & (k+\hat{k}_i)e^{i(k-\hat{k}_i)b_i}\end{pmatrix}\begin{pmatrix}E_i\\F_i\end{pmatrix}. \tag{59}$$

Putting

$$M_{a_i}=\frac{1}{2k}\begin{pmatrix}(k+\hat{k}_i)e^{-i(\hat{k}_i-k)a_i} & (k-\hat{k}_i)e^{i(k+\hat{k}_i)a_i}\\(k-\hat{k}_i)e^{-i(k+\hat{k}_i)a_i} & (k+\hat{k}_i)e^{i(\hat{k}_i-k)a_i}\end{pmatrix}, \tag{60}$$

and

$$M_{b_i}=\frac{1}{2\hat{k}_i}\begin{pmatrix}(k+\hat{k}_i)e^{-i(k-\hat{k}_i)b_i} & (\hat{k}_i-k)e^{i(k+\hat{k}_i)b_i}\\(\hat{k}_i-k)e^{-i(k+\hat{k}_i)b_i} & (k+\hat{k}_i)e^{i(k-\hat{k}_i)b_i}\end{pmatrix}, \tag{61}$$

and defining a 2×2 matrix M of element m_{ij}, such that

$$M=\prod_{i=0}^{n}(M_{a_i}\cdot M_{b_i}), \tag{62}$$

it is found that

$$\begin{pmatrix}A_0\\B_0\end{pmatrix}=M\begin{pmatrix}E_n\\0\end{pmatrix}. \tag{63}$$

The general coefficients in matrix form are then, for the reflection, $R=B_0/A_0$

$$R=\frac{m_{21}}{m_{11}}, \tag{64}$$

and for the transmission, $T=E_n/A_0$

$$T=\frac{1}{m_{11}}. \tag{65}$$

Determination of A_0 by considering unit-sinusoidal input volume velocity (a Dirac δ function in frequency) at the acoustic source, *defined now at* $x=0$, leads to a Green's function solution, $G_f(x|0|\omega)$, for the tract impulse response at some point beyond a terminating barrier at $x=l$, as follows.

On the assumption of a high-impedance, strictly plane-wave source, the glottal volume velocity $u_g'(0,t)$ is, from the Euler equation

$$u_g'(0,t)=-\frac{S(0)^{1/2}}{\rho_0}\int\frac{\partial\Psi(0,t)}{\partial x}dt. \tag{66}$$

The effects of a gradient in the boundary layer are not investigated in the present work but are of interest in detailed vocal-tract simulations that may also involve a time-varying impedance.[24]

Setting a unit sinusoidal input volume velocity, $u_g'(0,t)=e^{iwt}$, then

$$e^{iwt}=\frac{S(0)^{1/2}}{\rho_0 c}(A_0e^{iwt}-B_0e^{iwt}), \tag{67}$$

and hence

$$A_0=\frac{\rho_0 c}{S(0)^{1/2}(1-R)}. \tag{68}$$

The Green's function solution beyond a terminating barrier of width Δ at $x=l$ is, since $F_n=0$ and $E_n=A_0T$

$$G_f(x|0|\omega)=A_0Te^{-ik[x-(l+\Delta)]}. \tag{69}$$

Substitution of (68) into (69) gives

$$G_f(x|0|\omega)=\left(\frac{T}{1-R}\right)e^{-ik[x-(l+\Delta)]}, \tag{70}$$

where $\rho_0 c/S(0)^{1/2}$ has been normalized for convenience. In terms of the matrix coefficients

$$G_f(x|0|\omega)=\left(\frac{1}{m_{11}-m_{21}}\right)e^{-ik[x-(l+\Delta)]}. \tag{71}$$

Subsequently, the power spectrum is given by

$$|G_f(x|0|\omega)|^2=\left|\frac{1}{(m_{11}-m_{21})}\right|^2. \tag{72}$$

The complex Green's function may of course be expressed in modulus and argument form. From the previous discussion, it is clear, for example, that

$$\arg G_f(x|0|\omega)=-k[x-(l+\Delta)]+\arg T-\arg(1-R). \tag{73}$$

Resonance occurs for minima of $(1-R)$, that is, for R directed along the positive real axis, $\arg R=-2n\pi$, $n=0,1,2,...$, and spectral minima for $\arg R=-(2n+1)\pi$. Rearrangement of Eq. (26) gives

$$R=\frac{(S(0)Z_0/\rho_0 c)-1}{(S(0)Z_0/\rho_0 c)+1}, \tag{74}$$

and it is then straightforward to show that the resonance conditions are met for $X_0=0$, as must be the case.

It can be shown upon substitution of Eq. (74) into (70) that at resonance, the power spectrum is given by

$$|G_f(x|0|\omega)|^2=R_0, \tag{75}$$

as must be the case for a constant-velocity (current) source of unit amplitude.

The time-varying tract impulse-response, $G_f(x,t)$, may then be expressed as the finite Laplace inversion[25] of Eq. (71), and its convolution with an approximation to the volume velocity found at the human glottis, for example, will produce speech-like waveforms for appropriate potential functions.

For a single potential in general, the reflection $(x<a)$ and transmission $(x>b)$ coefficients are found from Eqs. (64) and (65) as

$$R=e^{-2ika}\frac{(k^2-\hat{k}^2)(1-e^{-2i\hat{k}\Delta})}{(k+\hat{k})^2-(k-\hat{k})^2e^{-2i\hat{k}\Delta}}, \tag{76}$$

and

$$T=\frac{4k\hat{k}e^{i(k-\hat{k})\Delta}}{(k+\hat{k})^2-(k-\hat{k})^2e^{-2i\hat{k}\Delta}}. \tag{77}$$

It can readily be shown that, as required for conservation of energy

$$|R|^2=1-|T|^2. \tag{78}$$

The Green's function formalism developed above is now demonstrated with the special case of a single square barrier.

1. $k^2<U_0$

Defining $\hat{k}\equiv -ik'$, as for the infinite step, the reflection coefficient may be written simply as $R=B/A$, with

$$R=\frac{U_0e^{-2ika}(e^{-2k'\Delta}-1)}{(2k^2-U_0)(e^{-2k'\Delta}-1)+2ikk'(e^{-2k'\Delta}+1)}, \tag{79}$$

and similarly for the transmission coefficient, $T=E/A$

$$T=\frac{4ikk'e^{(ik-k')\Delta}}{(2k^2-U_0)(e^{-2k'\Delta}-1)+2ikk'(e^{-2k'\Delta}+1)}. \tag{80}$$

Taking moduli and arguments of these coefficients leads, respectively, to a definition of the amplitude and phase shift (relative to incidence) of the reflected and transmitted waves. Thus, the reflection coefficient is found as

$$|R|^2=\left[1+\frac{4k^2(U_0-k^2)}{U_0^2\sinh^2 k'\Delta}\right]^{-1}, \tag{81}$$

and

$$\arg R=-2ak+\theta, \tag{82}$$

where

$$\theta=\arctan\left[\frac{2kk'\coth k'\Delta}{U_0-2k^2}\right]. \tag{83}$$

Similarly, the transmission coefficient is found as

$$|T|^2=\left[1+\frac{U_0^2\sinh^2 k'\Delta}{4k^2(U_0-k^2)}\right]^{-1}, \tag{84}$$

and

$$\arg T=k\Delta+\phi, \tag{85}$$

where

$$\phi=\arctan\left[\frac{(U_0-2k^2)}{2kk'\coth k'\Delta}\right], \tag{86}$$

such that

$$\tan\theta\tan\phi=1. \tag{87}$$

In the low-frequency limit, $\theta\to\pi$ and $\phi\to\pi/2$, in agreement with the Huygens–Fresnel principle of diffraction.

Before going on to apply these analytic results to acoustical simulations, the singularity in the plane-wave potential function at the duct termination must be addressed. As noted by Benade and Jansson, the curvature of a planar wavefront becomes infinite at the junction of the duct with free space, due to the discontinuity in area function. It is shown below, however, that *finite* barrier parameters may be fixed in comparison to known radiation impedance approximations, specifically that of the piston-in-infinite-baffle. The tract response to a time-varying source can then be obtained from Eq. (71), dealing elegantly with the problem of nonhomogeneous boundary conditions noted previously by Bonder. In this case, Sturm–Liouville analysis is not strictly appropriate, since it deals only with self-adjoint problems. The Green's function for operators and/or boundary conditions which are not self-adjoint can be expressed in terms of a singular value decomposition, sometimes referred to as a biorthogonal series decomposition[26] (although not factored explicitly in this way in the present paper).

B. Development of a finite barrier equivalent to a radiation impedance

In this section, it is shown that a one-dimensional barrier approximation may be made to the radiation impedance due

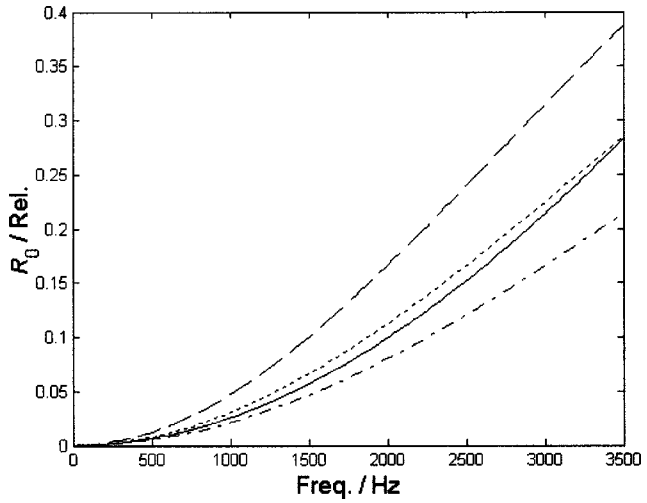


FIG. 8. R_0. Infinite baffle (———); $U_0 = 0.8\times10^5\ \mathrm{m}^{-2}$ (--), $1.0\times10^5\ \mathrm{m}^{-2}$ (···), $1.2\times10^5\ \mathrm{m}^{-2}$ (-·-).

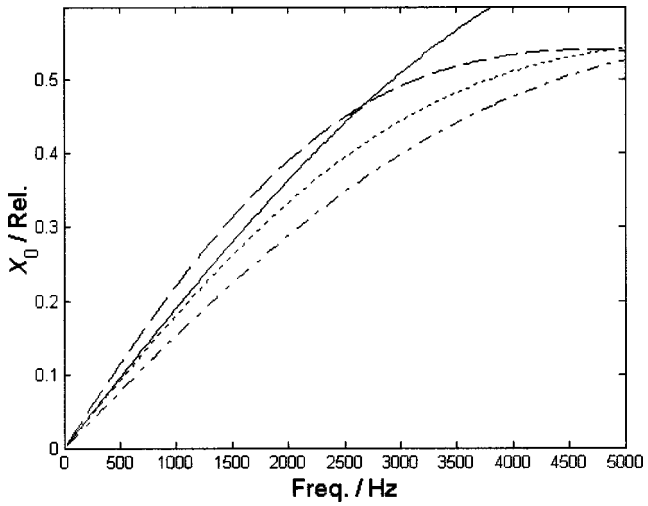


FIG. 9. X_0. Infinite baffle (———); $U_0 = 0.8\times10^5\ \mathrm{m}^{-2}$ (--), $1.0\times10^5\ \mathrm{m}^{-2}$ (···), $1.2\times10^5\ \mathrm{m}^{-2}$ (-·-).

to an infinite baffle. Such an approximation enables the efficient time-domain simulation of the resonating tract, either through finite-difference solution of Eq. (5) or the inverse Laplace transform of Eq. (71).

It is straightforward to show that, for a catenoidal section, the specific input impedance of a single potential barrier of arbitrary height and width is given by

$$Z_0 = \rho_0 c \frac{k}{\hat{k}} \left\{ \frac{1 + ik(\tan \hat{k}\Delta)/k}{k/\hat{k} + i \tan \hat{k}\Delta} \right\}, \tag{88}$$

and, for $k^2 \ll U_0$, may be approximated to first order as

$$Z_0 = \rho_0 c \left\{ \frac{k^2}{(U_0\Delta)^2} + i \frac{k}{U_0\Delta} \right\}. \tag{89}$$

In agreement with the principle of conservation of energy, the real part is inversely proportional to the surface area of radiation into the infinite cone, as can be seen by expanding Eq. (18) to first order. In the low-frequency limit, the term $(U_0\Delta)^2$ can be identified with the value 2π of the infinite baffle [Eq. (30)], appropriate to the solid angle of radiation into a half space. Calculations show that the agreement is reasonable only up to around 50 Hz, so that, for vowel simulations necessitating frequencies of up to at least 3000 Hz, terms beyond minimum order must be considered in defining a barrier approximation to the radiation impedance.

Derived from Eq. (88), the impedances associated with barriers of 1 mm width and height $(1.2,1.0,0.8)\times10^5\ \mathrm{m}^{-2}$ are illustrated in Figs. 8 and 9. Also shown are values appropriate to the infinite baffle,[27] on the assumption of a typical vocal-tract cross section of 5 cm^2. It can be concluded that, in the low-frequency range below around 3500 Hz, a barrier of height $1.0\times10^5\ \mathrm{m}^{-2}$ has comparable characteristics, particularly for the acoustic resistance. Figure 10 shows that the impedance due to this ''end barrier'' obeys the limiting conditions $R_0 \to 1$ and $X_0 \to 0$, and avoids the ripple due to the infinite baffle (small numerical spikes are noted), presumably due to the underlying continuity in the area function.

The associated coefficients of wave function reflection ($|R|^2$) and transmission ($|T|^2$) are plotted in Fig. 11. Transmission is ~75% at the top of the plane-wave range for the 1-mm barrier, falling to around 10% on widening to 4 mm. The end barrier, then, fulfills the prediction made in Sec. III B, namely, that substantial wave function transmission, or wave-mechanical ''tunneling,'' can be expected for barriers of width less than the wave-mechanical penetration depth of around 3 mm. It is noted that this methodology avoids the constraints inherent in the WKB approximation to the transmission as proposed by Benade and Jansson, constraints pointed out again recently by Kergomard (personal communication 2001), specifically that the potential function varies slowly over a wavelength and that there is no backward-propagating solution within the barrier region.

The Green's-function methodology was then applied to the end barrier in the configuration of Fig. 12, for $l = 17.5$ cm. Equation (72) yielded the spectrum of Fig. 13, with formants at 472, 1422, 2382, 3352, and 4330 Hz, that is, at approximately (94–96)% of the nominal values, comparable to those due to the infinite baffle, and with a similar

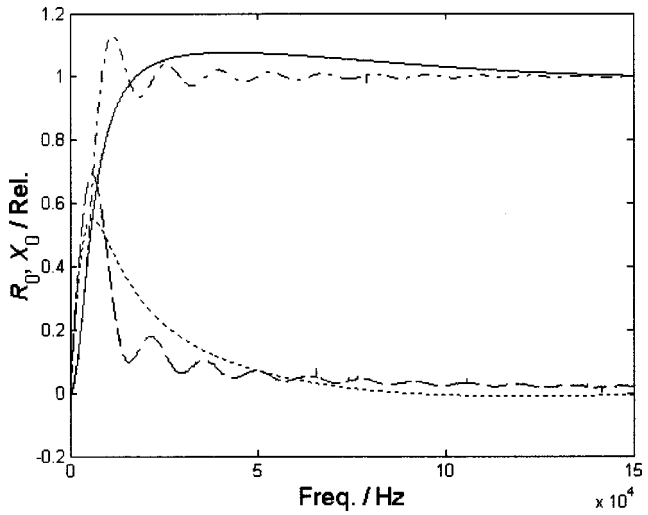


FIG. 10. Infinite baffle, R_0 (-·-) and X_0 (--); compared to end barrier, R_0 (———), and X_0 (···), up to high-frequency limit.

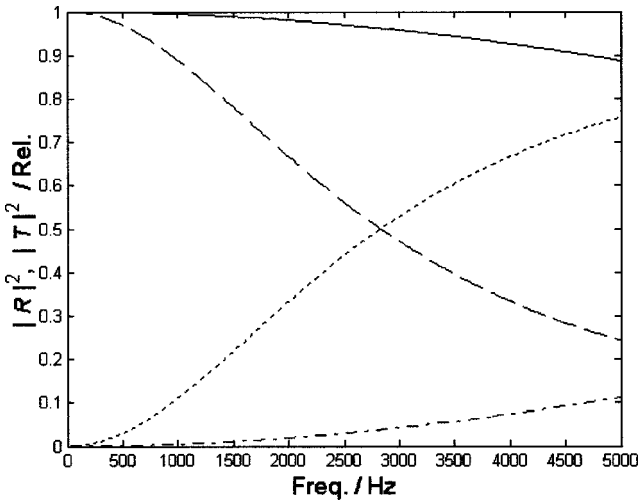


FIG. 11. $\Delta=1$ mm: $|R|^2$ (--), and $|T|^2$ ($\cdots$); $\Delta=4$ mm: $|R|^2$ (———), and $|T|^2$ (-·-).

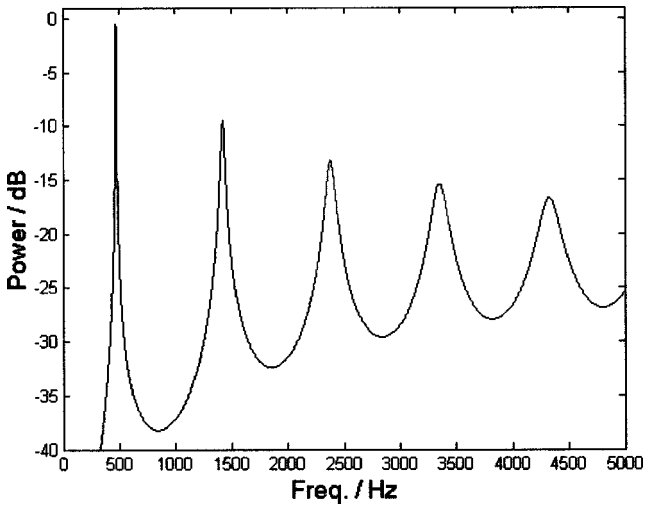


FIG. 13. Green's function power spectrum for the duct system of Fig. 12, for $l=17.5$ cm.

spectral slope of around 6 dB oct^{-1} in the low-frequency range.

An equivalence set of area functions is shown in Fig. 14, where the only constraint on initial conditions is that, as discussed in Sec. II for the vocal tract, $S(x)$ remain approximately within the range (0.2–15) cm^2. The matching of straight or conical segments to an infinite cone, by a 1-mm-wide hyberbolic section at $l=17.5$ cm according to Eq. (15), is illustrated for a few possible pairs of initial conditions on $S(0)^{1/2}$. Although the angle of infinite conicity varies, the radius at distances comparable to the shortest wavelengths is always several orders of magnitude above that in the duct itself. The compression of the area-function information into the stark potential-function profile of Fig. 12 is quite striking. As noted in Sec. II, the dispersive properties of the set are identical; it is now clear that, since they share a wave-mechanical Green's function, their relative resonance characteristics are also identical.

To confirm this point, the time-varying tract impulse response, $G_f(x,t)$, was obtained from the finite Laplace inversion of Eq. (71) and compared to the finite difference solution of both the Klein–Gordon equation for the potential-function profile of Fig. 12, and the Webster equation for three arbitrary, equivalent area functions. Solutions were compared at $L=l+\Delta$. As shown in Fig. 15, the analytic and numerical results are identical within the accuracy of the finite difference method. Further, they were found to be identical, up to a scaling factor of $S(L)$, with those of the equivalence set. The potential-function formalism with a barrier equivalent to a radiation impedance is therefore being investigated in time-domain vowel simulations.[12]

1. $k^2 > U_0$

Since the plane-wave singularity at the duct termination must be approximated as a high, thin barrier, a detailed wave-mechanical treatment of propagation above the barrier will not be given. However, phenomena analogous to the "transmission resonances" in nuclear physics, for example, will occur. Low, and perhaps long, barriers may indeed be relevant in other acoustical applications, and the transfer matrix and Green's function analysis followed exactly.

In this section, a fully analytic Green's-function methodology has been developed that extends readily to arbitrary strings of square potential functions, and it has been found

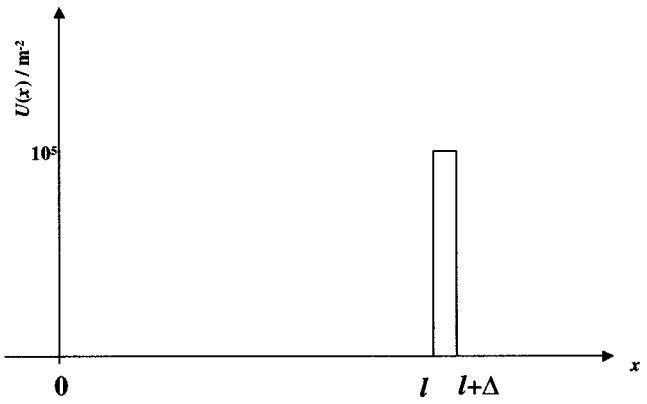


FIG. 12. Potential-function profile of duct terminated in end barrier, $U_0=1.0\times10^5\ \text{m}^{-2}$, $\Delta=1$ mm.

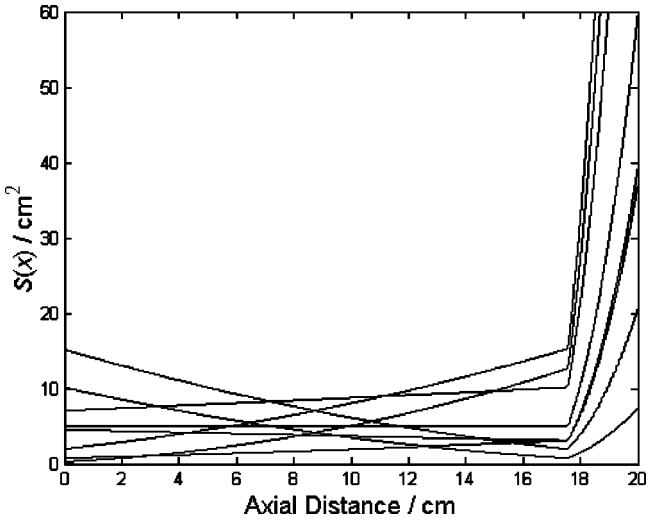


FIG. 14. Equivalence set of area functions for the duct system of Fig. 12, for $l=17.5$ cm.

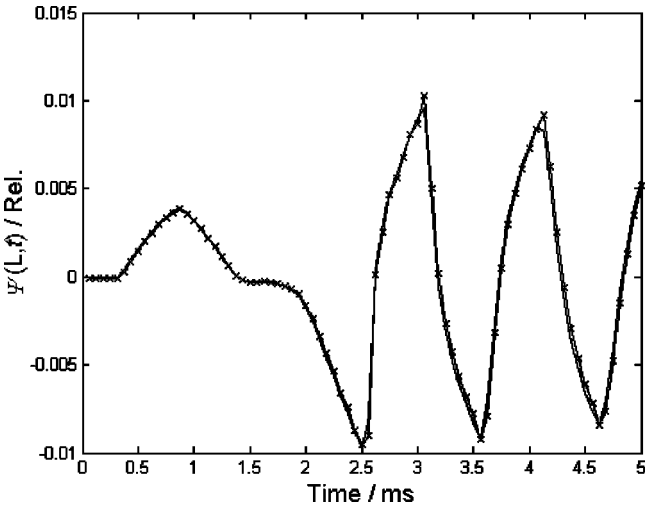


FIG. 15. Inverse Laplace transform, $G_f(\mathrm{L},t)$ ($\cdots\times\cdots$), finite difference numerical solution (———).

that the singularity at the duct termination can be approximated by a square barrier of finite dimensions. In paper II of this series, we will go on to discuss constricted tract geometries as described by potential wells and well–barrier pairs, representing highly localized trigonometric contractions and hyperbolic expansions.

V. CONCLUSIONS

It has been noted that one-dimensional duct acoustics can be theoretically described by an equation familiar from relativistic quantum theory, the Klein–Gordon equation. Solutions of the acoustical Klein–Gordon equation have been discussed with particular reference to the dispersions due to the positive, piecewise-constant, potential functions, namely, the infinite step and finite barrier. It has been shown that many-to-one mappings exist between any given potential function and duct area functions, and an equivalence set of area functions for a single potential barrier has been found. Since one-dimensional cylindrical- and conical-segment models fail to capture dispersive phenomena, and refer directly to the area function, it is concluded that the wave-mechanical model of an acoustical tract is both more accurate and more compact than either. The work is relevant to any area of numerical modeling or discretization of acoustical waveguides and the compact description of ducts, including speech parametrization applications and acoustic reflectometry.

ACKNOWLEDGMENTS

The authors thank the Engineering and Physical Sciences Research Council of the United Kingdom for their support of this work, which was primarily carried out in the Department of Physics, King's College London, under an EPSRC studentship. It was completed in the Department of Environmental and Mechanical Engineering, The Open University, UK, under EPSRC Grant No. GR/N32471/01. The manuscript has benefited greatly from the comments of an anonymous referee, to whom the authors extend their thanks. Useful discussions were also held with Dr. Shahram Taherzadeh of the Open University.

[1] E. Eisner, "Complete solutions of the 'Webster' horn equation," J. Acoust. Soc. Am. **41**, 1126–1146 (1967).
[2] A. G. Webster, "Acoustical impedance, and the theory of horns and the phonograph," Proc. Natl. Acad. Sci. U.S.A. **5**, 275–282 (1919).
[3] V. Salmon, "Generalized plane wave horn theory," J. Acoust. Soc. Am. **17**, 199–211 (1946).
[4] A. H. Benade and E. V. Jansson, "On plane and spherical horns with nonuniform flare. I. Theory of radiation, resonance frequencies, and mode conversion," Acustica **31**, 79–98 (1974).
[5] N. Amir, E. Rosenhouse, and V. Shimony, "Input impedance of musical horns and the 'horn function'," Appl. Acoust. **38**, 15–35 (1993).
[6] D. P. Berners and J. O. Smith, "On the use of Schrödinger's equation in the analytic determination of horn reflectance," Int. Comp. Music Conf. Proc. 419–422 (1994).
[7] D. P. Berners, "Acoustics and signal processing techniques for physical modelling of brass instruments," Ph.D. thesis, Department of Electrical Engineering, Stanford University, CA (1999).
[8] D. H. Keefe and A. Barjau, "Acoustic propagation in flaring, axisymmetric horns. II. Numerical results, WKB theory, and viscothermal effects," Acustica **85**, 285–293 (1999).
[9] B. J. Forbes, "A potential-function analysis of speech acoustics," Ph.D. thesis, Department of Physics, King's College London, Strand, London WC2R 2LS (2000).
[10] B. J. Forbes and E. R. Pike, "The time-independent Klein–Gordon equation: A perturbation analysis," Paper presented at the 144th Meeting of the Acoustical Society of America, Cancun, Mexico (2–6, Dec. 2002).
[11] B. J. Forbes, "The vocal-tract potential function: Many-to-one mappings and compensatory articulations in speech production," Poster presented at the 144th Meeting of the Acoustical Society of America, Cancun, Mexico (2–6, Dec. 2002).
[12] B. J. Forbes and E. R. Pike, "A potential-function parametrization of vowel acoustics," Paper (invited) presented at DAGA'03, Aachen, Germany (18–21, March 2003) (in press).
[13] B. J. Forbes, D. B. Sharp, J. A. Kemp, and A. Li, "Singular system methods in acoustic pulse reflectometry," Acustica (to be published).
[14] V. Pagneux, N. Amir, and J. Kergomard, "A study of wave propagation in varying cross-section waveguides by modal decomposition. I. Theory and validation," J. Acoust. Soc. Am. **100**(4), 2034–2048 (1996).
[15] A. M. Bruneau, M. Bruneau, P. H. Herzog, and J. Kergomard, "Boundary layer attenuation of higher order modes in waveguides," J. Sound and Vibration. **119**(1), 15–27 (1987).
[16] L. Eriksson, "Higher order mode effects in circular ducts and expansion chambers," J. Acoust. Soc. Am. **68**, 545–550 (1980).
[17] R. W. Pyle, Jr., "Notes on one-parameter horn theory," Work in progress, received by private communication (2001).
[18] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. I, Chap. 2, p. 139.
[19] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. II, Chap. 9, p. 1092.
[20] N. Amir, G. Rosenhouse, and U. Shimony, "A discrete model for tubular acoustic systems with varying cross section—the direct and inverse problems. I. Theory," Acustica **81**, 450–462 (1995).
[21] M. Kac, "Can one hear the shape of a drum?" Am. Math. Monthly **73**, 1–23 (1966).
[22] J.-P. Dalmont, C. J. Nederveen, and N. Joly, "Radiation impedance of tubes with different flanges: Numerical and experimental investigations," J. Sound Vib. **244**(3), 505–534 (2001).
[23] L. Bonder, "The *n*-tube formula and some of its consequences," Acustica **52**, 216–226 (1983).
[24] M. Kob, "Physical Modelling of the Singing Voice," Ph.D. thesis, University of Technology, Aachen, Germany. (Logos, Berlin, 2002).
[25] P. M. Morse and H. Feshbach, in Ref. 19, Chap. 11, p. 1340.
[26] P. M. Morse and H. Feshbach, in Ref. 18, Chap. 7, p. 884.
[27] W. E. Zorumski, "Generalized radiation impedances and reflection coefficients of circular and annular ducts," J. Acoust. Soc. Am. **54**(6), 1667–1673 (1973).

Cure monitoring using ultrasonic guided waves in wires

T. Vogt, M. Lowe, and P. Cawley
Department of Mechanical Engineering, Imperial College, London SW7 2BX, United Kingdom

(Received 15 May 2002; revised 24 April 2003; accepted 5 May 2003)

The possibility of using ultrasonic guided waves for monitoring the cure process of epoxy resins is investigated. The two techniques presented use a wire waveguide which is partly embedded in the resin. The first technique is based on the measurement of attenuation due to leakage of bulk waves into the resin surrounding the waveguide. The second technique measures the reflection of the guided wave that occurs at the point where the waveguide enters the resin. Both the attenuation and the reflection coefficient change significantly during cure, and the numerical methods to relate these to the material properties of the curing resin are presented in this paper. The results from the modeling are experimentally verified and show good agreement. The applicability of each testing method is discussed, and typical cure-monitoring curves are presented. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1589751]

PACS numbers: 43.20.Mv, 43.35.Ze [DEC]

I. INTRODUCTION

There is a need in the automotive industry for a technique which allows one to monitor the cure process of adhesive joints in hostile industrial environments.[1] Controlling the cure process of adhesive joints in large structures can be difficult because of temperature variations in the ovens and the sensitivity of the cure process to temperature. One approach currently being investigated is to embed a thin wire waveguide into the bond line and to monitor changes in the ultrasonic propagation characteristics of a guided wave. The results from such measurements could be used to assess and improve the quality of the adhesive bonding.

There is also a scientific interest in an additional measurement technique for monitoring the cure process of epoxy resins in general. The results from the wire-waveguide technique can be used to gain more insight into the physical process of cure. Especially useful would be the comparison with other cure-monitoring techniques such as ultrasonic bulk wave techniques,[2–9] and dielectric analysis.[10,11]

Several researchers have previously investigated the possibility of using a waveguide technique for cure monitoring. In 1974 Papadakis presented two methods for determining the change in material properties of an epoxy resin over the cure cycle.[12] The first method is based on the measurement of the attenuation of longitudinal and torsional guided waves in a circular bar. The second method measures the reflection of those waves from the point where the waveguide enters the resin. Both attenuation and reflection change significantly during cure. The feasibility of using guided waves in wires for monitoring cure as well as the structural integrity of composites was also investigated by Harrold and Sanjana.[13] Their measurements were based on the attenuation method. Li *et al.* used the wire-waveguide technique in a through-transmission arrangement, with the transmitter at one end of the wire and the receiver at the other end.[14] Where it is embedded in the epoxy, the wire is bent in the form of a U-turn, so that two sections of the wire run parallel to each other. When the epoxy cures, the guided wave attenuation will change. At one point, the attenuation becomes very high, and the energy is not transmitted entirely along the wire but also between the two parallel legs in the embedded section. Characteristic curing curves can be obtained this way.

However, the results so far are only of a qualitative nature and have not been related to the actual material properties of the epoxy resin. The current investigation revisits the two methods presented by Papadakis, and the aim of this paper is to give the measurement techniques a theoretical foundation based on guided wave theory and numerical modeling. The physical insight gained from this analysis enables the design of guided wave systems to the specific needs of each technique. In the following, some general features of guided waves are introduced, and the experimental setup is presented. Then, the reflection coefficient and attenuation methods are explained in detail, and experiments are presented to validate the modeling. Finally, a discussion deals with the implications of the findings on the applicability of each method.

II. GUIDED WAVES IN WIRES

The cure-monitoring techniques described in this paper use ultrasonic waves in wire waveguides. For the following analysis, it is necessary to introduce some underlying principles of guided wave propagation. A cylindrical coordinate system (r,Θ,z) is chosen, with the axis of the cylindrical waveguide in the z direction.

A guided wave is characterized by its phase velocity, c_{ph}, and attenuation coefficient, α. In general, these quantities depend on the frequency f, and this relationship is referred to as dispersion. A guided wave can be thought of as a superposition of partial plane waves which are reflected inside the waveguide. In a free waveguide surrounded by vacuum, the guided wave is attenuated due to intrinsic losses of the waveguide material only. This makes long-range transmission in low loss materials like metals possible. In a perfectly elastic free wire, the attenuation of the guided wave is zero, and the wave is fully characterized by its phase velocity dispersion curve. Once the guided wave is embedded in an-

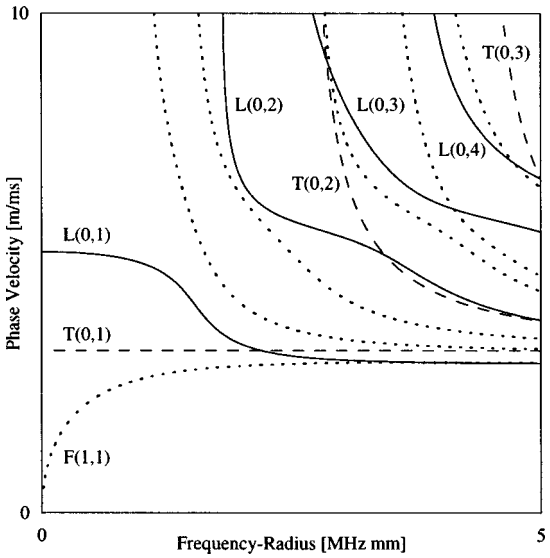


FIG. 1. Phase velocity dispersion curves for a steel wire in vacuum. Longitudinal modes are plotted in solid lines (———), torsional modes in dashed lines (--), and flexural modes in dotted lines (···). For clarity, higher-order flexural modes are not labeled.

other material, the guided wave is attenuated due to the leakage of bulk waves into the surrounding medium. This is explained in detail in Sec. V. Its phase velocity may change too, due to mass loading.

Three different families of modes are present in wire waveguides: longitudinal, torsional, and flexural modes. The cure-monitoring techniques presented in this paper concentrate on the lowest order longitudinal mode, L(0,1). However, explanations will also be given for the lowest order torsional mode, T(0,1), since this mode was also included in the analysis by Papadakis.[12] The mode-naming convention in this paper follows the one given by Silk and Bainton.[15] In this notation, the first number indicates the circumferential order, and, being zero for both L(0,1) and T(0,1), denotes that these modes are both axially symmetric. The reason for considering fundamental modes at low frequencies becomes apparent when considering the phase velocity dispersion curves for a steel wire (see Fig. 1). First, at higher frequencies many modes are present, which can interfere with each other and make the experimental analysis complicated. Second, it is found that most modes show dispersion. Therefore, a pulsed signal which propagates along the wire does not retain its shape, which can also complicate the analysis. However, T(0,1) does not show any phase velocity dispersion at all, and L(0,1) exhibits only little dispersion at low frequencies. This makes them favorable for testing purposes. One other property of the curves to note is that they scale with the product of frequency and the radius of the wire. Thus, the horizontal axis which is used in Fig. 1 is the frequency–radius product.

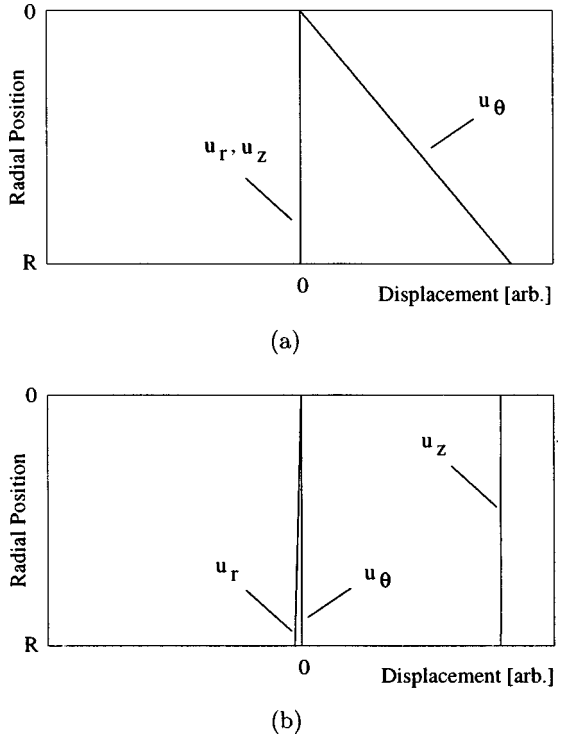


FIG. 2. Displacement mode shapes of (a) T(0,1) and (b) L(0,1) at 0.1 MHz mm. *R* denotes the outer radius of the wire.

Every mode has a characteristic radial distribution function, referred to as the mode shape. Consider, for example, the components of the displacement **u**. Figure 2(a) shows the displacement mode shape of T(0,1) at 0.1 MHz mm in a circular steel wire. It exhibits shear displacements in the angular direction only, and propagates at the bulk shear velocity of the waveguide. The mode shape of L(0,1) at this frequency–radius product, however, is dominated by the axial displacements; there is only little displacement in the radial direction, and no displacement in the angular direction [see Fig. 2(b)]. We refer to the L(0,1) mode as ''longitudinal,'' but the term ''extensional'' is also often used in the literature for this mode in slender waveguides.

III. EXPERIMENTAL SETUP

The general setup for a cure-monitoring experiment is shown in Fig. 3. For low frequencies, it is convenient to excite longitudinal and torsional waves by means of magnetostrictive transduction. A section of the magnetostrictive waveguide is permanently magnetized. An alternating current through a coil, which acts as a transducer and is placed over the magnetized section, provides an additional time-varying axial magnetic field. Thereby, longitudinal modes are created in the wire if it is axially magnetized (Joule effect), and torsional waves are created if it is magnetized in

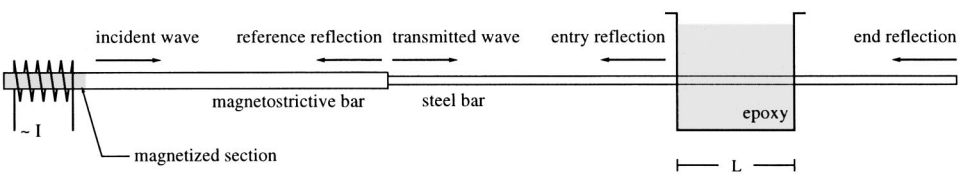


FIG. 3. Experimental setup. A magnetostrictive device is used to excite the longitudinal mode, L(0,1), and the torsional mode, T(0,1), respectively. Alternatively, a piezoelectric device is used for the excitation of L(0,1) at higher frequencies.

the circumferential direction (Wiedemann effect).[16] The inverse magnetostrictive effect allows for the same coil to be used as a receiver in a pulse–echo arrangement. A commonly used magnetostrictive material is the alloy Remendur (see, for example, Ref. 17).

The magnetostrictive wire is connected to a steel wire, which can be chosen to have a different wire diameter and acts as the measurement wire. The connection between the magnetostrictive and the measurement wire was accomplished by drilling a hole approximately 1–2 mm deep into the larger diameter wire at one end, with a diameter matching that of the smaller diameter wire. The smaller diameter wire is then glued into the hole using instant glue. The guided wave propagates along the magnetostrictive wire until it is partially reflected at the interface to the steel wire. This reflection serves as a reference to cancel out effects such as changes in the excitation amplitude. The fraction of the guided wave which is transmitted into the measurement wire propagates further until it interacts with the embedding epoxy resin.

Alternatively, a piezoelectric transduction mechanism can be used for the excitation of the longitudinal mode, L(0,1). In this case, the magnetostrictive wire is replaced by a steel bar, and a piezoelectric disk is bonded to one end. An alternating voltage applied to the electrodes of the disk creates axial displacements. An additional steel backing is used to generate a reaction and thus a significant force in the wire, and the transducer, consisting of the piezoelectric element and the steel backing, is operated below its resonance frequency so as to avoid distortion of the input pulse shape. Even though the backing limits the upper operating frequency, depending on the dimensions of the disk and the backing, this arrangement was found to be better suited for higher frequencies (above around 0.4 MHz) than the magnetostrictive transducer.

Once the steel wire is embedded in the resin, the reflection coefficient and the attenuation of the guided wave are calculated with reference to the case when the steel wire is in vacuum (see Secs. IV and V, respectively). Therefore, the reflection from the end of the wire is recorded before embedding the wire. Note that the attenuation due to leakage when the waveguide is embedded in air is negligible, and is therefore practically equivalent to the case where the waveguide is in vacuum. Also, it has to be ensured that the end of the wire is as square as possible in order to avoid mode conversion, particularly into the fundamental flexural mode, F(1,1).

As explained in Sec. IV, a reflection of considerable amplitude may occur at the point where the wire enters the resin. For the reflection coefficient method, this reflection is monitored, and it is important that the dimensions of the sample and the embedded length of the wire are chosen such that there is no interference with other reflections, for example from the end of the wire. When using the attenuation technique, the reflection from the end of the wire is monitored, so it is important that the entry reflection is small. The details concerning each individual setup are explained in later sections.

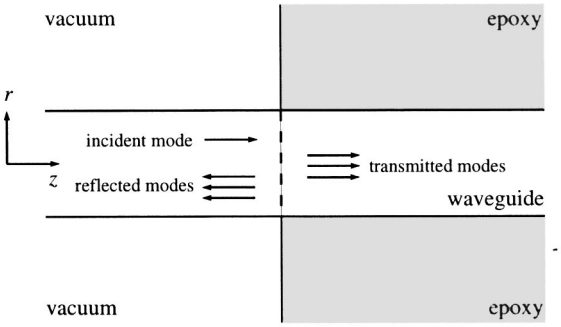


FIG. 4. Schematic illustrating the scattering process at the point where a free wire waveguide enters an embedding epoxy resin.

IV. REFLECTION METHOD

A. Background

Consider a guided wave which is incident from the free section of a wire (see Fig. 4). At the point where the wire enters the epoxy resin, a change in surface impedance occurs. The interface between the free and the embedded section acts as a scatterer. This means that, depending on the change in surface impedance, the incident guided wave is partly reflected into the modes of the free cylinder and transmitted as guided wave modes of the embedded cylinder.

Finite-element (FE) modeling using a time-marching algorithm can be used to characterize this entry reflection. A description can be found in Ref. 18. However, FE modeling is difficult, for example, for viscous liquids, where the shear velocity is low, because this results in a very small element size and time step. Therefore, this method is not practical for predictions in the early stages of epoxy cure. In an alternative approach, the scattering process can also be analyzed in terms of the modes which are involved and their mode shapes. We refer to this as the modal solution. The particle velocity and stress field in the free section can be thought of as a superposition of the normal modes of the infinitely long free wire, whereas the field in the embedded part can be thought of as a superposition of the leaky modes of the embedded wire. At the interface, the boundary conditions must be fulfilled, and this leads to a modal description of the scattering process. A detailed analysis is given in Ref. 18; therefore, only the main findings will be presented in this paper.

In the following sections, the nature of the entry reflection is explained in detail, and the results are verified experimentally.

B. Characteristics of the entry reflection

Both torsional and longitudinal modes are of circumferential order zero. Due to the axial symmetry of the scattering problem, it is evident that an incident axisymmetric mode can only be scattered into a mode which is also axisymmetric. Therefore, no flexural modes are excited at the interface. Also note that, for geometrical reasons, mode conversion between longitudinal and torsional modes does not occur. The testing frequencies are chosen to be lower than the cutoff frequencies of L(0,2) and T(0,2), respectively. Therefore, the fundamental modes L(0,1) and T(0,1) can only be re-

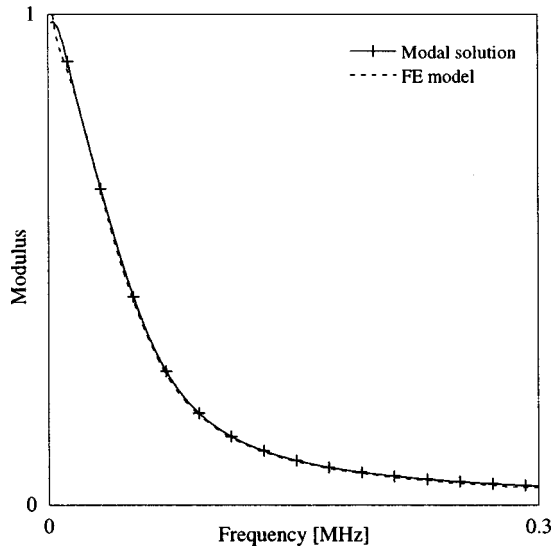


FIG. 5. Example of the entry reflection coefficient of L(0,1) for a 1-mm-radius steel wire which is partly embedded in an epoxy resin. The material properties are given in Table I.

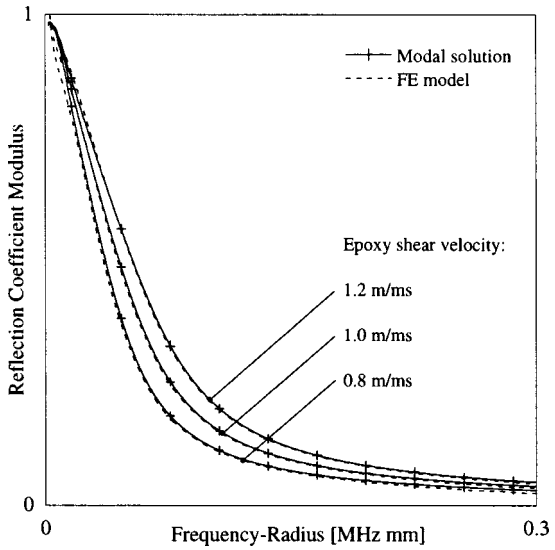


FIG. 6. Entry reflection coefficient of L(0,1) in a steel wire which is partly embedded in an epoxy resin for different epoxy shear velocities. All other material properties are as for solid epoxy in Table I.

flected into themselves and into nonpropagating and inhomogeneous modes of the free cylinder. However, the latter modes are local vibrations which do not propagate energy away from the interface and whose fields are attenuated very strongly.[19] This simplifies the scattering problem, since in an experiment, where the receiving transducer is relatively far away from the scattering region, the only modes received are the reflected L(0,1) or T(0,1) mode, depending on which mode is generated by the transmitter. The analysis carried out in this paper and the one presented in Ref. 18 concentrates on the longitudinal mode L(0,1). However, the analysis can easily be repeated for the torsional mode, T(0,1).

Figure 5 shows an example of FE and modal predictions for the entry reflection of L(0,1) in a 1-mm-radius steel wire which is partly embedded in a solid epoxy resin. The materials are modeled as purely elastic, and a summary of the material parameters used for the analysis is given in Table I. The mode shapes of the guided wave system modeled here scale with the product of frequency and radius, and this is also true for the reflection coefficient. This was confirmed by modeling different wire diameters in the range between 0.5 and 2 mm. It can be seen that the reflection is high for low frequency–radius products, and falls off to low values at higher frequency–radius products. It has also been found that the L(0,1) entry reflection is almost independent of the longitudinal bulk velocity of the embedding material.[18]

For a cure-monitoring experiment it is important to know how the reflection coefficient changes with the increasing shear velocity of the embedding resin. Figure 6 shows the entry reflection of L(0,1) for different epoxy shear velocities. Apart from the shear velocity, which is indicated in the plot for each reflection coefficient curve, the epoxy properties are the same as the ones given in Table I for the solid epoxy. It should be noted that the density is assumed to stay constant throughout cure. While the epoxy resin solidifies, the shear velocity increases and the difference in surface impedance between the free and the embedded section becomes larger, so that the entry reflection increases. Figure 7 shows the reflection coefficient as a function of the epoxy shear velocity at a frequency–radius product of 0.05 MHz mm. Of course, the longitudinal bulk velocity also undergoes significant changes during cure, but as mentioned above, this does not affect the reflection coefficient.

As will be explained in Sec. VI, the shear bulk attenua-

TABLE I. Material parameters used for the analysis.

Material	Long. bulk velocity [m/s]	Shear bulk velocity [m/s]	Density [kg/m^3]
Steel	5960	3260	7932
Solid epoxy	2610	1000	1170
Liquid epoxy	1700	···	1170

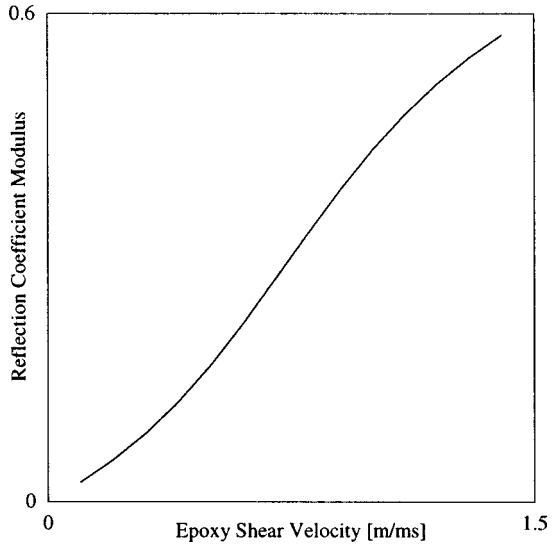


FIG. 7. Predicted entry reflection coefficient of L(0,1) as a function of the epoxy shear velocity at a frequency–radius product of 0.05 MHz mm (modal solution). All other material properties are as for solid epoxy in Table I.

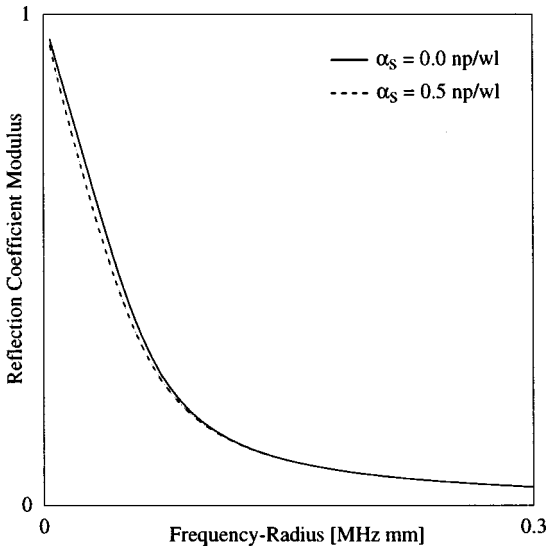


FIG. 8. Predicted entry reflection coefficient of L(0,1) for different bulk shear attenuation values of the embedding epoxy resin (modal solution). All other material properties are according to Table I for solid epoxy.

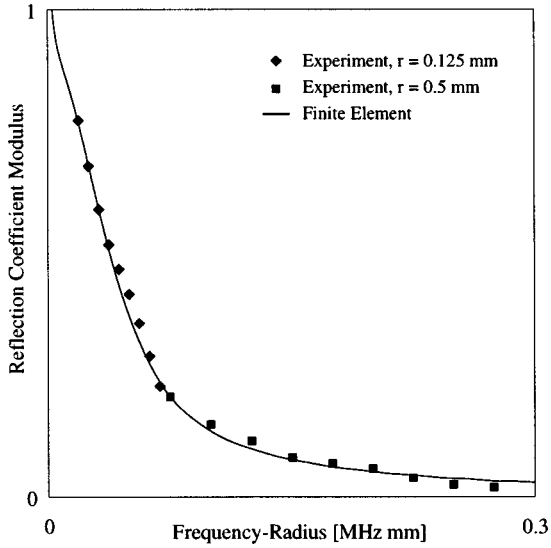


FIG. 10. Measured and predicted entry reflection coefficient for steel wires partly embedded in an epoxy resin. Experiments were conducted with two wires of different radius, $r=0.125$ mm and $r=0.5$ mm. The solid line shows the finite-element results.

tion decreases with the progress of cure, whereas the shear bulk velocity increases. For typical values of bulk shear velocity and attenuation in the epoxy after the completion of cure, results using the modal solution suggest that even high attenuation values in the epoxy influence the reflection coefficient only minimally (see Fig. 8). For this analysis, the frequency dependence of the bulk shear attenuation of the resin was assumed to be such that the attenuation per wavelength was constant. This behavior is supported by findings of Freemantle *et al.*[5] in ultrasonic bulk wave experiments.

C. Experimental verification of frequency dependence

For the experimental validation in this section, the influence of the shear bulk attenuation on the reflection coefficient after the resin has fully cured is neglected, so the resin is modeled as an elastic material. The accuracy of this approximation is discussed in Sec. VI.

Figure 9(a) shows an example of an experimental time trace of L(0,1) in the free wire before embedding. It shows the reference reflection and the reflection from the end of the wire. The time trace after embedding is shown in Fig. 9(b), with the resin fully cured. Here, the entry reflection is received at an earlier time because of the reduction in propagation distance.

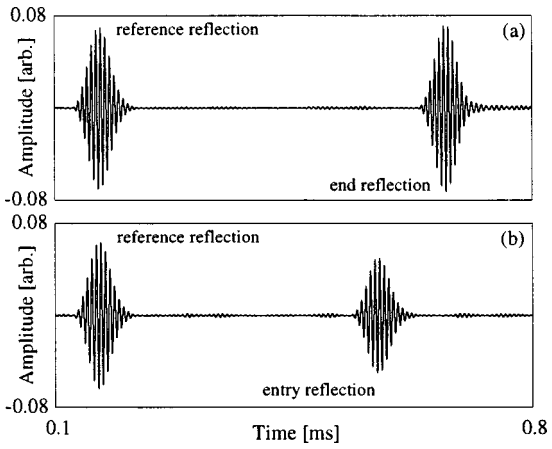


FIG. 9. Experimental time traces of the lowest order longitudinal mode, L(0,1), in a 0.125-mm-radius steel wire. Plots (a) and (b) show the time trace before and after embedding in epoxy resin, respectively. In plot (b) the resin has fully cured.

Since piezoelectric transducers were used for these experiments, a narrow-band signal had to be chosen in order to prevent signal distortion due to the excitation of frequencies close to the resonance frequency of the transducer. Let the modulus of the Fourier transform of the reference reflection in the free and the embedded wire be denoted by A_0^{vac} and A_0^{emb}, respectively, the end reflection in the free wire by A^{vac}, and the entry reflection in the embedded wire by A^{emb}. Then, the reflection coefficient R is calculated according to

$$R=\frac{A^{\mathrm{emb}}/A_0^{\mathrm{emb}}}{A^{\mathrm{vac}}/A_0^{\mathrm{vac}}}. \tag{1}$$

Experiments were conducted using the lowest order longitudinal mode, L(0,1). Two steel wires of different radius, $r=0.125$ mm and $r=0.5$ mm, were embedded in the same sample of epoxy resin (Araldite 2012, Ciba Specialty Chemicals) in order to confirm the frequency–radius dependence of the reflection coefficient. Measurements were taken between 0.1 and 0.55 MHz, and the reflection coefficient was evaluated at the center frequencies.

In a separate experiment, the material properties of the cured epoxy sample were measured using conventional ultrasonic transducers in through transmission at 0.05 MHz. The shear velocity was determined as $c_s=0.885$ m/ms, and the density $\rho=1170$ kg/m^3. Using these material properties, the reflection coefficient was predicted by means of FE modeling, assuming the epoxy to be perfectly elastic. The experimental results are plotted together with the predictions in Fig. 10. The agreement is very good, and it confirms the frequency–radius dependence of the reflection coefficient.

V. ATTENUATION METHOD

A. Background

It is clear from Sec. IV that at higher frequency–radius products the guided wave incident from the free section of the wire is almost entirely transmitted into the embedded section. This is an important result and its implications will be discussed in detail in Sec. VI. The analysis in this section deals with the guided wave which is transmitted as a mode in the embedded waveguide.

As mentioned before, a guided wave propagates without attenuation in a free elastic waveguide since the partial waves are fully reflected inside the waveguide. The wave is then fully characterized by the phase velocity and the frequency. However, if the waveguide is embedded in another material, the partial waves are not only reflected from the waveguide boundaries, but may also be transmitted. The guided wave is therefore attenuated due to the leakage of bulk waves into the surrounding medium. If the embedding material is an ideal liquid, only bulk longitudinal waves are radiated, whereas if it is a solid, both shear and longitudinal waves may be radiated and contribute to the leakage.

The attenuation rate depends on both the material properties of the waveguide and the surrounding medium. In general, leakage occurs when the phase velocity of the guided wave is larger than the lowest bulk velocity of the embedding material (in general the shear velocity). If the guided wave phase velocity is lower than this, the guided wave becomes nonleaky, in which case the fields in the embedding material decrease exponentially with distance from the interface, and the energy remains trapped close to the surface of the waveguide. A well-known example of such an interface wave is the Scholte wave (see, for example, Ref. 20). However, due to the nature of the problem presented here, we will only deal with leaky waves. In this case, it is found that the larger the difference in the characteristic acoustic impedance between the waveguide and the embedding material, the lower is the attenuation.

Since the leakage of longitudinal bulk waves is a function of the radial surface displacements of the waveguide and thus is zero for torsional waves, the attenuation of T(0,1) is exclusively due to shear leakage. The radial displacements of L(0,1) at low frequency–radius products are small compared to the axial displacements, so shear leakage also controls the attenuation of L(0,1) in this range. Therefore, these modes are particularly sensitive to the shear properties of the embedding material.

B. Modeling of attenuation

Like the phase velocity, the attenuation coefficient (or simply attenuation, for brevity) is a property of the guided wave and forms part of the solution to the dispersion equation of the particular guided wave system under consideration. The multipurpose software tool DISPERSE has been used[21] to trace the velocity and attenuation dispersion curves and to calculate the mode shapes.

As an example, Fig. 11 shows the attenuation dispersion curves of L(0,1) for a steel wire embedded in a liquid epoxy resin, modeled as an ideal liquid and representing the precure state, and a solid epoxy resin, modeled as an elastic solid and representing the postcure state. The density is assumed to stay constant throughout cure. As mentioned before, for an ideal liquid epoxy, only longitudinal bulk waves radiate, whereas for the solid epoxy, both shear and longitudinal waves contribute to the leakage. The dispersion curve for the liquid epoxy case confirms that at low frequency–radius products, the attenuation due to the leakage of longitudinal waves is small. Therefore, the attenuation is mainly sensitive to the shear properties of the epoxy resin. It should also be noted that the ordinate of Fig. 11 is an np·mm/m-scale. Therefore, the attenuation per meter of a 1-mm-radius wire at a given frequency–radius product is double that of a 2-mm-radius wire.

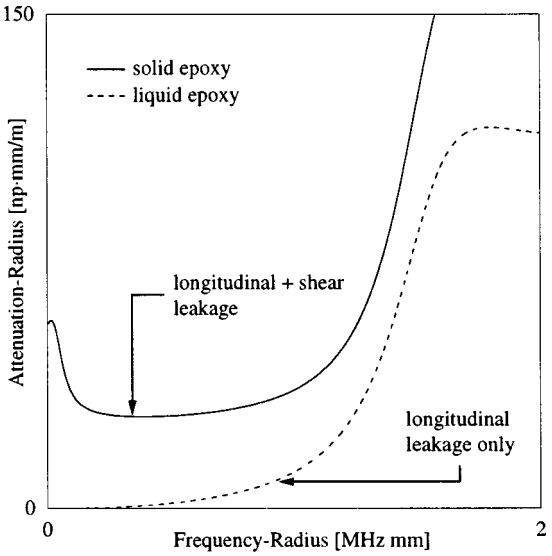


FIG. 11. Predicted attenuation dispersion curves of L(0,1) for a steel wire embedded in a liquid and a solid epoxy. The material properties are given in Table I.

Figure 12 shows the effect of different epoxy longitudinal velocities on the L(0,1) attenuation for a cured epoxy and an imaginary uncured one with a low shear velocity and the same longitudinal velocity. It is found that the longitudinal velocity has only little influence in the low frequency–radius limit below 0.5 MHz mm after the early stages of the cure cycle. Only at higher frequency–radius products does the longitudinal velocity have a significant effect on the attenuation due to the increasing radial displacements at the surface of the waveguide. This is similar to the behavior observed for the entry reflection.

The effect of shear bulk attenuation on the attenuation of the longitudinal guided wave at the end of cure is considered in Fig. 13. This shows the attenuation of L(0,1) for a purely elastic and a strongly attenuating epoxy resin. As before, for the evaluation of the influence of bulk shear attenuation on the reflection coefficient, the attenuation was taken to be constant per wavelength, thus linearly increasing with frequency. It can be seen that the difference in guided wave attenuation is very small, and may therefore also be neglected when calculating the shear properties of the epoxy resin after cure.

The L(0,1) attenuation as a function of the epoxy shear

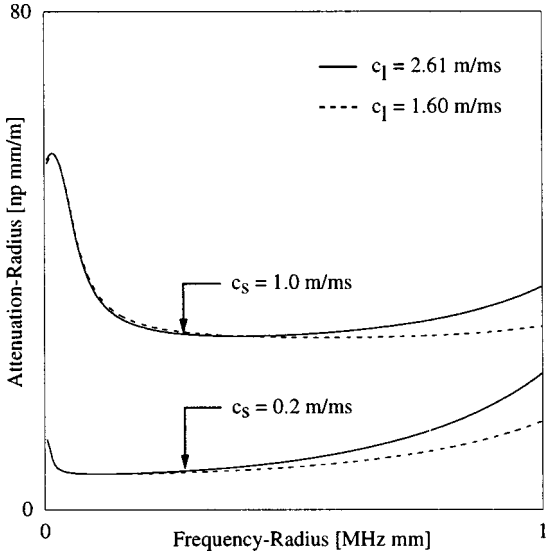


FIG. 12. Predicted attenuation dispersion curves of L(0,1) in a steel wire for different epoxy longitudinal velocities for a solid epoxy ($c_s = 1.0$ m/ms) and an imaginary uncured epoxy resin with low shear velocity ($c_s = 0.2$ m/ms). The density of the epoxy resin in this analysis is $\rho = 1170$ kg/m^3, and the steel properties are given in Table I.

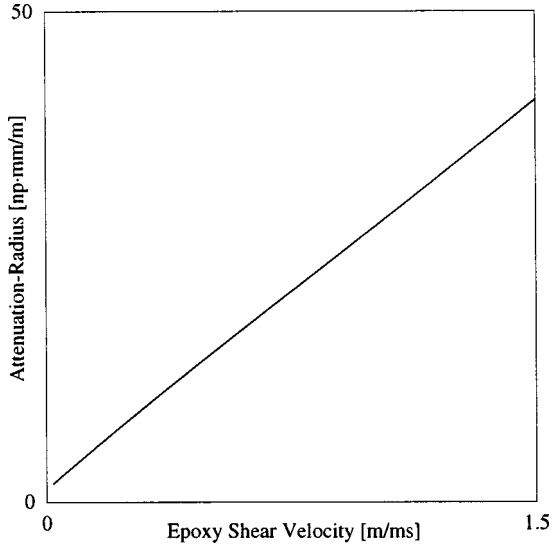


FIG. 14. Predicted dependence of the L(0,1) attenuation as a function of the solid epoxy shear velocity at 0.4 MHz mm.

velocity is given in Fig. 14, assuming that the epoxy resin is elastic. The attenuation increases almost linearly with increasing shear velocity. This is to be expected, since by increasing the epoxy shear velocity, the difference between the acoustic impedance of the waveguide and the epoxy resin decreases.

It has been shown in previous publications that the attenuation of guided torsional and longitudinal waves can be used to measure the material properties of an embedding material. Mostly it has been applied to the measurement of the viscosity, density, and longitudinal velocity of liquids.[22–25] In this case the entry reflection is negligible, since the shear velocity introduced by the shear viscosity is small. However, one needs to be careful when the embedding material is solid, since the entry reflection may be significant. This is discussed below.

To calculate the attenuation from experimental data, let the modulus of the Fourier transform of the reference reflection in the free and the embedded wire be denoted by A_0^{vac} and A_0^{emb}, respectively, and the end reflection in the free and the embedded wire by A^{vac} and A^{emb}, respectively. If L is the length of the embedded section of the wire, the attenuation coefficient is calculated in a pulse–echo setup in the following way:

$$\alpha = -\frac{1}{2L}\ln\left(\frac{A^{\mathrm{emb}}/A_0^{\mathrm{emb}}}{A^{\mathrm{vac}}/A_0^{\mathrm{vac}}}\right). \tag{2}$$

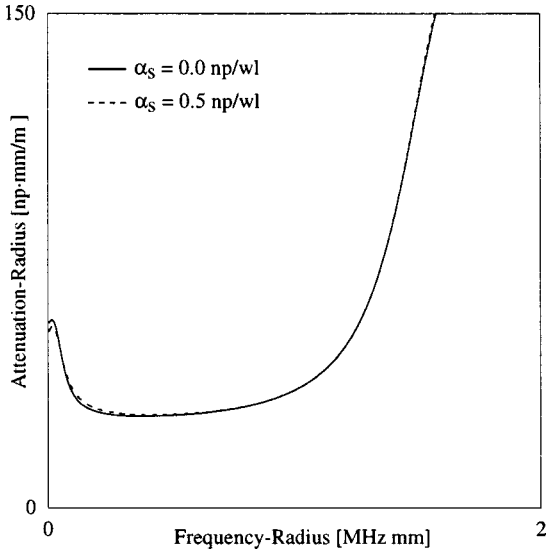


FIG. 13. Predicted attenuation dispersion curves of L(0,1) for a solid epoxy (see Table I) without attenuation (solid line) and with a constant attenuation of 0.5 np/wl (dashed line).

VI. APPLICABILITY OF REFLECTION AND ATTENUATION METHOD

The nature of the entry reflection coefficient of the fundamental longitudinal mode determines the range of frequency–radius products at which each method is applicable. Clearly, the reflection coefficient method can only be used at low frequency–radius products, since only there is it of significant magnitude. For the attenuation measurements, it would be preferable to also work at low frequency–radius products, since the influence of the longitudinal properties of the adhesive can be neglected there, which makes it possible to relate the attenuation measurement to the shear properties of the epoxy resin. Also, the longitudinal mode becomes dispersive at higher frequencies, which makes testing more difficult. However, in practice, where the exciting transducer is placed on a free section of the waveguide, and the guided wave of the free bar has to be transmitted into the guided wave in the embedded section, the need to minimize entry reflection forces one to work at higher frequency–radius products. Only there is enough energy transmitted into the embedded part, and the entry reflection may be neglected or

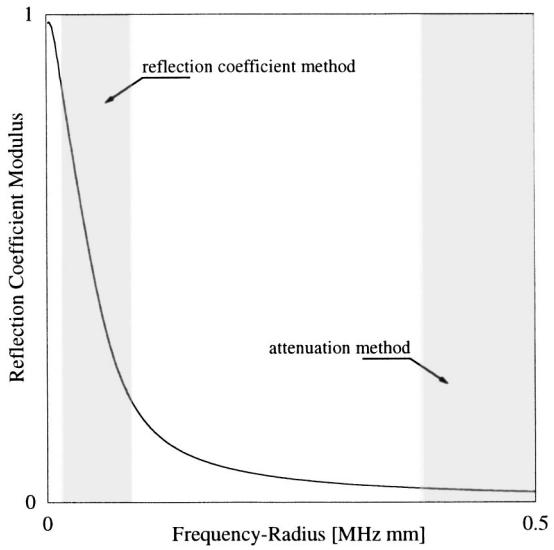


FIG. 15. Entry reflection coefficient of L(0,1) as plotted in Fig. 5. The shaded areas indicate schematically where the reflection and attenuation method are applicable for cure-monitoring experiments.

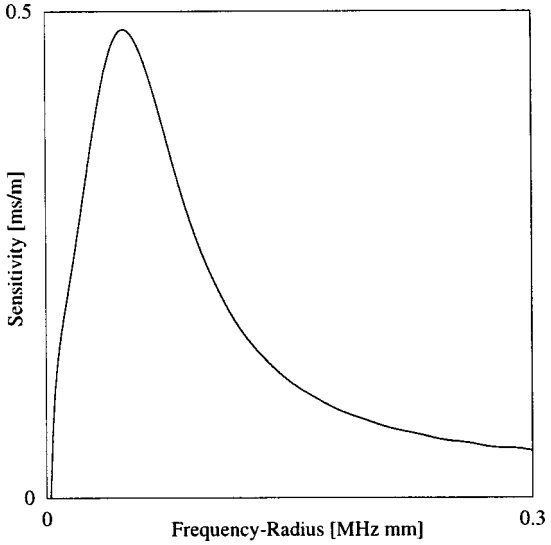


FIG. 16. Sensitivity of the L(0,1)-reflection coefficient to changes in the shear velocity at $c_s = 1.0$ m/ms. A maximum occurs at around 0.05 MHz mm.

included in the analysis as an error. The shaded regions in Fig. 15 schematically indicate when each method is applicable. Of course, the limits of these regions are not strict, and they depend on the system under investigation and the accuracy required.

The shear velocity of an epoxy resin changes significantly during cure. At the beginning of cure, the material properties range from viscous to pasty, depending on the epoxy resin. At this stage, conventional ultrasonic bulk wave techniques fail to monitor changes in the shear properties due to the high shear bulk attenuation. One significant advantage of the wire-waveguide method is that the shear properties can be measured from the start over the whole curing time. During solidification the shear velocity increases and the shear attenuation decreases steadily due to the cross linking of polymer chains.[5] These changes happen relatively quickly at the beginning of cure, whereas towards the end of cure, the curing rate is slowed down very much, even though the cure cycle is not complete.

It should be noted that the modeling examples given in Secs. IV and V are idealized in that the epoxy resin was modeled as an elastic material. As mentioned before, in reality it behaves more like a visco-elastic material with a changing bulk shear velocity and attenuation. Since only one variable can be determined either from the reflection coefficient or from the attenuation technique, it is not possible to characterize the embedding material completely. However, it has been shown that in the later stages of cure, the bulk shear attenuation has a negligible effect, so the bulk shear velocity can be predicted accurately. In the earlier stages of cure, the bulk shear attenuation is very high and will affect the calculation of the shear velocity. Therefore, an accurate value of shear velocity cannot be obtained. Nevertheless, the measurement is indicative of the progress of cure in the early stages.

VII. RESULTS

A. Reflection method

The modeling tools presented enable the optimization of the guided wave system, by choice of the frequency–radius product, to have the highest sensitivity to shear changes. The sensitivity S, which is the change in reflection coefficient R for a given change in shear velocity c_s at a particular shear velocity c_0, is determined by the following expression:

$$S = \left.\frac{\mathrm{d}R}{\mathrm{d}c_s}\right|_{c_0}. \tag{3}$$

Figure 16 shows the sensitivity around a shear velocity of $c_0 = 1$ m/ms. For a typical epoxy resin surrounding a steel wire, a maximum occurs at around 0.05 MHz mm. The reflection coefficient experiments were therefore conducted in this frequency–radius range.

The commercial epoxy adhesive Araldite 2013 (Ciba Specialty Chemicals) was used in the experiments. It is a low-temperature curing adhesive for metal-bonding and gap-filling applications. Cure-monitoring experiments were conducted using a 0.5-mm-radius steel wire as the measurement wire. The longitudinal mode L(0,1) was excited in the wire using magnetostrictive transduction. A wideband, two-cycle Hanning-windowed signal was chosen so that the reflection coefficient could be evaluated at several frequencies. The epoxy was cured at room temperature in a bulk sample of 50-ml volume in a polyethylene beaker of 53-mm inner diameter, and the measurement wire was fed through holes in the beaker. After taking a reference time trace before embedding, the adhesive was poured into the beaker and measurements were taken at intervals of 5 min. The beaker was positioned approximately in the middle of the measurement wire.

Figure 17 shows example time traces obtained during cure. The time trace at 0 h shows the reference time trace before embedding. The signal denoted by ''A'' is the refer-

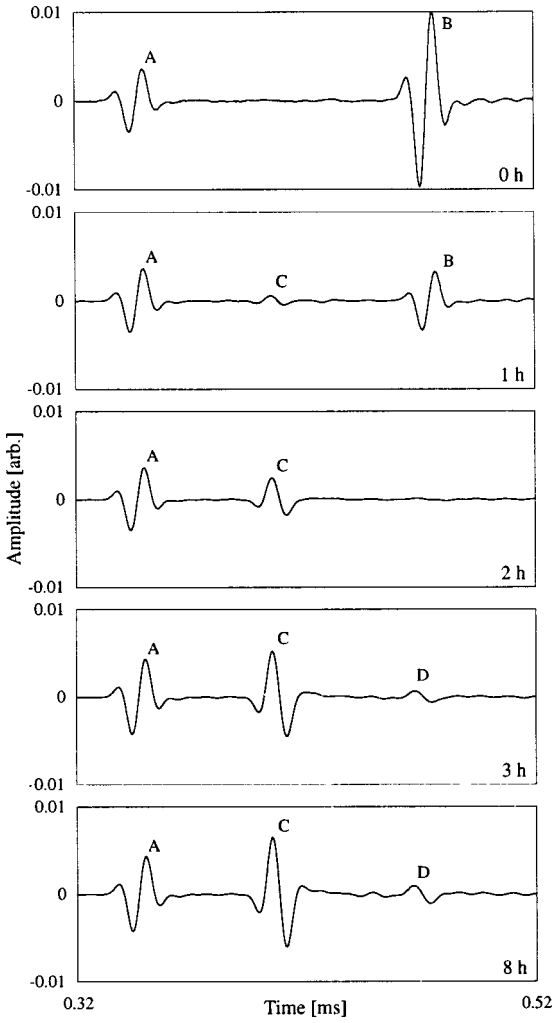


FIG. 17. Time traces during a reflection coefficient cure measurement. The plot at 0 h shows the time trace in the free wire before embedding.

ence reflection from the joint between the magnetostrictive and the measurement wire, and the reflection denoted by "B" is the reflection from the end of the wire. The remaining time traces show the reflections while the measurement wire is embedded in the curing epoxy. After 1 h, the entry reflection, "C," can be seen, and the amplitude of the end reflection decreases due to increasing attenuation in the embedded part and also due to the fact that some of the energy is reflected at the entry point. After 2 h, the end reflection disappears completely, and only the entry reflection remains visible. The entry reflection increases monotonically with curing time due to the increasing shear velocity in the epoxy. A second reflection, "D," can be observed after approximately 3 h. This is not a reflection from the end of the wire but a reverberation of the entry reflection "C" between the entry point and the joint between the magnetostrictive and the measurement wire. No significant further changes are observed after approximately 8 h.

The cure-monitoring curves obtained from this at different frequencies are plotted in Fig. 18, using Eq. (1). As expected, the reflection coefficient is higher at lower frequencies. The reflection coefficient of L(0,1) at the end of cure at 0.05 MHz mm was $R=0.451$. With a density of $\rho=1170$ kg/m^3, this corresponds to an epoxy shear velocity of c_s $=1.06$ m/ms (see Fig. 7). This agrees well with the shear velocity data given in Ref. 5.

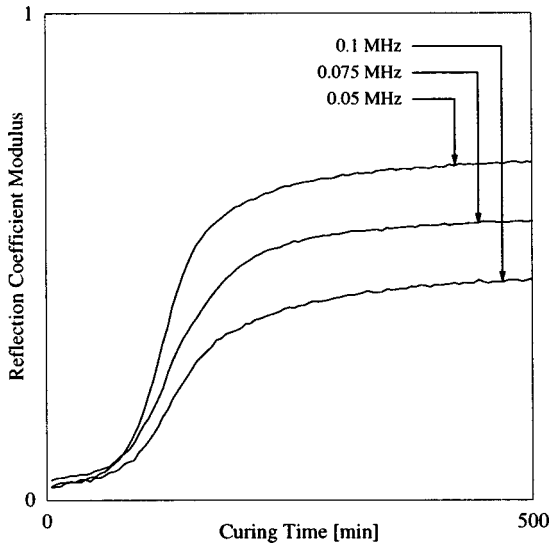


FIG. 18. Cure-monitoring curves obtained in Araldite 2013 with the reflection coefficient method at different frequencies.

Note that in this case cure is measured at the edge rather than in the interior of the epoxy sample. This is important to keep in mind as considerable temperature differences may occur between the inside and the surface of the curing adhesive, resulting in different curing rates.

B. Attenuation method

The setup was changed slightly in order to make cure-monitoring measurements using the attenuation method. As pointed out earlier, these measurements have to be taken at a higher frequency–radius product in order to avoid reflection at the entry point. Therefore, measurements were taken using a 1.5-mm-radius steel wire at 0.266 MHz, which corresponds to a frequency–radius product of about 0.4 MHz mm. A piezoelectric transducer was used to excite L(0,1) in a 1-mm-radius steel wire, and in order to avoid ringing a ten-cycle Hanning-windowed input signal was applied. The embedding length L for the calculation of the guided wave attenuation was 53 mm. The same adhesive was used as for the reflection coefficient experiments.

Figure 19 shows the development of the time traces during cure. The reference signal from the joint between the steel wires is denoted by "A," and the reflection from the end of the measurement wire is denoted by "B." The plot at 0 h shows the reference time trace before embedding. After 1 h, the amplitude of the end reflection has decreased due to the fact that the wire is embedded and leaking bulk waves into the epoxy. As the polymer chains are cross linking, the shear velocity increases and the amplitude decreases monotonically until the end of cure as the L(0,1) attenuation increases. After 2 h, a small entry reflection occurs, "C," which increases until the end of cure. This entry reflection reaches a value of approximately $R=0.05$ by the end of cure.

The attenuation as a function of the curing time is plotted in Fig. 20. In order to check consistency between the reflection coefficient and the attenuation results, the epoxy

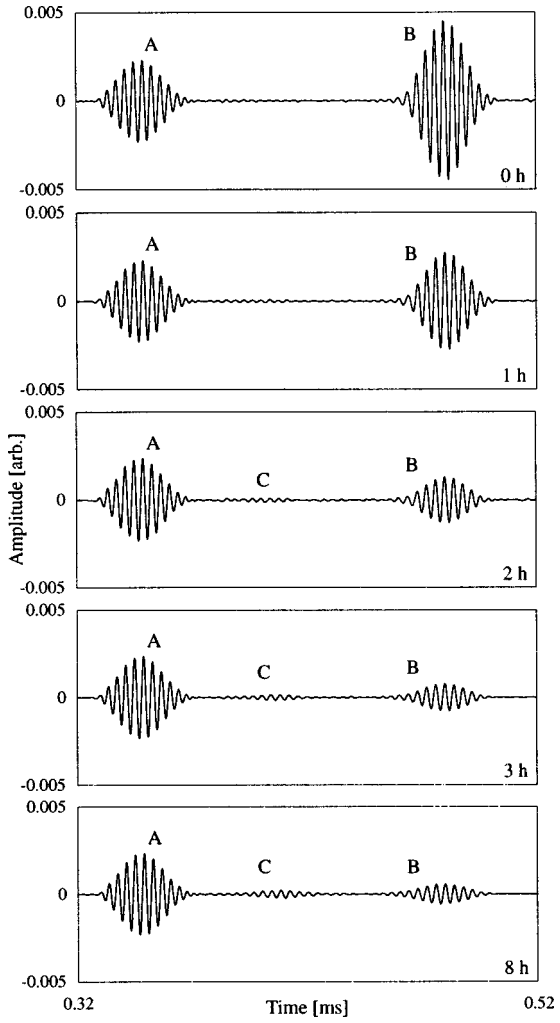


FIG. 19. Time traces during an attenuation cure measurement. The plot at 0 h shows the time trace in the free wire before embedding.

shear velocity is calculated at the end of cure. The attenuation of L(0,1) in the 1.5-mm-radius steel wire at the end of cure was $\alpha=18.89$ np/m, which corresponds to an attenuation–radius product of $\alpha\cdot r=28.33\,\mathrm{np\cdot mm/m}$. Using these values, the epoxy shear velocity is calculated as $c_s=1.01\,\mathrm{m/ms}$. This is within 4% of the value obtained with the reflection coefficient method, and also agrees well with the shear velocity data given by Freemantle *et al.*[5]

It should be noted that the entry reflection always introduces an error in the attenuation measurements if it is not accounted for in the analysis; the measured attenuation appears larger than the actual attenuation. The only way that this can be avoided is to fully embed the waveguide in the epoxy resin. If this is not possible, an estimate of the maximum error introduced by the entry reflection can be calculated from the entry reflection which was also measured during the cure experiment (reflection "C"). The transmission coefficient T may be calculated from the energy balance equation

$$R^2+T^2=1. \tag{4}$$

One main assumption enters this equation. It is assumed that all the incident power is scattered into the transmitted and reflected L(0,1) modes, and not into other modes. This would

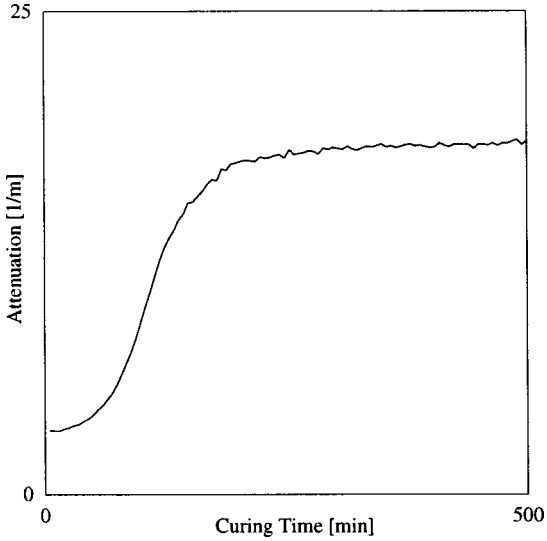


FIG. 20. Cure-monitoring curve obtained in Araldite 2013 with the attenuation method at 0.4 MHz mm.

seem reasonable if, apart from the propagating L(0,1) modes in the free and the embedded part, there are only nonpropagating and inhomogeneous modes present at the measurement frequency. This is true in the free bar, but in the embedded section, the corresponding modes are not necessarily nonpropagating or inhomogeneous. However, this has been discussed in Ref. 18, and it was found that their influence on the reflection and transmission coefficient is minimal. Thus, the above equation is fulfilled to a high degree of accuracy.

In order to calculate the error introduced by the entry reflection, it is further assumed that the transmission coefficient from the free section into the embedded section is equal to the transmission coefficient from the embedded into the free section. It should be noted that this is not necessarily true, and the difficulty encountered when trying to calculate the latter transmission coefficient is outlined in Ref. 18. In a pulse–echo experiment, as schematically shown in Fig. 3, the end of the wire waveguide protrudes from the epoxy sample in order to avoid losses through the end face of the wire. The received end reflection was therefore transmitted twice from the free into the embedded section and twice from the embedded into the free section, each time with a transmission coefficient of T. Thus, the normalized amplitude $A^{\mathrm{emb}}/A_0^{\mathrm{emb}}$ can be written in the following way:

$$\frac{A^{\mathrm{emb}}}{A_0^{\mathrm{emb}}}=\frac{A^{\mathrm{vac}}}{A_0^{\mathrm{vac}}}T^4e^{-2\alpha L}. \tag{5}$$

It follows that the error ε introduced when calculating the attenuation according to Eq. (2) instead of Eq. (5) is

$$\varepsilon=\frac{2}{L}\ln T. \tag{6}$$

Note that this error becomes smaller with increasing embedding length L. Since T is always smaller than 1, $\varepsilon<0$, and therefore the actual attenuation is overestimated. In this case, with $R=0.05$, the error is estimated as $\varepsilon\approx-0.05$ np/m. This is small compared to the measured attenuation and may be neglected.

In contrast to the reflection coefficient method, the measured shear velocity for the attenuation method presents an average value measured over the embedded length of the waveguide.

VIII. CONCLUSIONS

Two cure-monitoring methods based on ultrasonic guided waves in wires have been presented, and their theoretical foundations have been established. The first method is based on the measurement of the reflection which occurs where a waveguide enters the curing material, whereas the second method measures the guided wave attenuation over the embedded length. The main advantage of the wire-waveguide technique is that it enables the measurement of shear properties throughout the cure process, including the period just after the onset of cure, which is not accessible by means of bulk ultrasonic techniques.

It was shown that at low frequencies both methods are sensitive to the shear properties, and are only minimally dependent on the longitudinal properties. Both reflection coefficient and attenuation increase with the shear velocity of the curing material, and characteristic monitoring curves have been obtained. The reflection coefficient method can only be applied at relatively low frequency–radius products, whereas the attenuation method is only applicable at relatively high frequency–radius products. Using the modeling tools presented in this paper, it is possible to design a guided wave system such that it is suited to work optimally for the desired cure-monitoring method. Cure-monitoring curves have been presented, and shear velocities at the end of cure have been consistently predicted by both techniques.

ACKNOWLEDGMENTS

This work is funded by the Engineering and Physical Science Research Council (EPSRC), and carried out in collaboration with Jaguar Cars Ltd., UK, and Ford Motor Company, UK.

[1] I. N. Moody, P. A. Fay, and G. D. Suthurst, *Can Adhesives Meet the Challenge of Vehicle Bodyshell Construction*? (Elsevier Applied Science, New York, 1986).
[2] G. A. Sofer and E. A. Hauser, ''Polymerization of thermosetting polymers,'' J. Polym. Sci. **8**, 611–620 (1952).
[3] J. H. Speake, R. G. C. Arridge, and G. J. Curtis, ''Measurement of the cure of resins by ultrasonic techniques,'' J. Phys. D **7**, 412–424 (1974).
[4] A. M. Lindrose, ''Ultrasonic wave and moduli changes in a curing epoxy resin,'' Exp. Mech. **18**, 227–232 (1978).
[5] R. J. Freemantle and R. E. Challis, ''Combined compression and shear wave ultrasonic measurements on curing adhesive,'' Meas. Sci. Technol. **9**, 1291–1302 (1998).
[6] S. I. Rohklin, D. K. Lewis, K. F. Graff, and L. Adler, ''Real-time study of frequency dependence of attenuation and velocity of ultrasonic waves during the curing reaction of epoxy resin,'' J. Acoust. Soc. Am. **79**, 1786–1793 (1986).
[7] R. E. Challis, R. P. Cocker, A. K. Holmes, and T. Alper, ''Viscoelasticity of thin adhesive layers as a function of cure and service temperature measured by a novel technique,'' J. Appl. Polym. Sci. **44**, 65–81 (1992).
[8] A. Maffezzoli, E. Quarta, V. A. M. Luprano, G. Montagna, and L. Nicolais, ''Cure monitoring of epoxy matrices for composites by ultrasonic wave propagation,'' J. Appl. Polym. Sci. **73**, 1969–1977 (1999).
[9] D. D. Shepard and K. R. Smith, ''A new ultrasonic measurement system for the cure monitoring of thermosetting resins and composites,'' J. Therm. Anal. **49**, 95–100 (1997).
[10] B. G. Martin, ''Monitoring the composite cure cycle by dielectric analysis,'' Mater. Eval. **34**, 49–54 (1976).
[11] F. I. Mopsik, S. S. Chang, and D. L. Hunston, ''Dielectric measurements for cure monitoring,'' Mater. Eval. **47**, 448–453 (1989).
[12] E. P. Papadakis, ''Monitoring the moduli of polymers with ultrasound,'' J. Appl. Phys. **45**, 1218–1222 (1974).
[13] R. T. Harrold and Z. N. Sanjana, ''Acoustic waveguide monitoring of the structural integrity of composite materials,'' Polym. Eng. Sci. **26**, 367–372 (1986).
[14] Y. Li, G. J. Posakony, and S. M. Menon, ''Monitoring the cure process of polymer composites and neat resin using an ultrasonic wire waveguide technology,'' in 43rd International SAMPE Symposium (1998), 937–948.
[15] M. G. Silk and K. F. Bainton, ''The propagation in metal tubing of ultrasonic wave modes equivalent to Lamb waves,'' Ultrasonics **17**, 11–19 (1979).
[16] N. S. Tzannes, ''Joule and Wiedemann effects—The simultaneous generation of longitudinal and torsional stress pulses in magnetostrictive materials,'' IEEE Trans. Sonics Ultrason. **SU-13**, 33–41 (1966).
[17] L. C. Lynnworth, *Ultrasonic Waveguide Experiments Booklet* (Panametrics, Inc., 1987).
[18] T. Vogt, M. J. S. Lowe, and P. Cawley, ''The scattering of guided waves in partly embedded cylindrical structures,'' J. Acoust. Soc. Am. **113**, 1258–1272 (2003).
[19] B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, FL, 1990), Vol. 2.
[20] J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge, UK, 1999).
[21] B. N. Pavlakovic, M. J. S. Lowe, D. N. Alleyne, and P. Cawley, ''DISPERSE: A general purpose program for creating dispersion curves,'' in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. 16, pp. 185–192.
[22] L. C. Lynnworth, *Ultrasonic Measurements for Process Control: Theory, Techniques, Applications* (Academic, San Diego, 1989).
[23] J. O. Kim, H. H. Bau, Y. Liu, L. C. Lynnworth, S. A. Lynnworth, K. A. Hall, S. A. Jacobson, J. A. Korba, R. J. Murphy, M. A. Strauch, and K. G. King, ''Torsional sensor applications in two-phase fluids,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 563–574 (1993).
[24] T. Vogt, M. Lowe, and P. Cawley, ''Ultrasonic waveguide techniques for the measurement of material properties,'' in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, New York, 2002), Vol. 21, pp. 1742–1749.
[25] C. L. Shepard, B. J. Burghard, M. A. Friesel, B. P. Hildebrand, X. Moua, A. A. Diaz, and C. W. Enderlin, ''Measurement of density and viscosity of one- and two-phase fluids with torsional waveguides,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 536–548 (1999).

Oscillatory limited compressible fluid flow induced by the radial motion of a thick-walled piezoelectric tube

Dong-Youn Shin[a)]
Manchester Materials Science Center, UMIST, Grosvenor Street, Manchester M1 7HS, United Kingdom

Paul Grassia
Department of Chemical Engineering, UMIST, P.O. Box 88, Manchester M60 1QD, United Kingdom

Brian Derby
Manchester Materials Science Center, UMIST, Grosvenor Street, Manchester M1 7HS, United Kingdom

(Received 2 August 2002; revised 20 June 2003; accepted 3 July 2003)

A simple oscillatory, slightly compressible, fluid flow model in a thick-walled piezoelectric tube used in a drop-on-demand inkjet print head is developed from the point of view of fluid-structure interaction to take account of pressure wave propagation and pressure loading opposing wall motion. A frequency sweep is performed computationally using the model revealing the first acoustic fluid-structure resonance frequency and the influence of fluid viscosity. The validity of the model, with given information on the speed of sound in a fluid, is evaluated by comparing the theoretically predicted resonance frequency to the experimentally measured resonance frequency. In addition, the intrinsic speed of sound can be easily computed using the measured acoustic resonance frequency and this computed speed of sound agrees closely with speeds of sound reported in the literature. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1603769]

PACS numbers: 43.20.Tb, 43.20.Mv [MO]

I. INTRODUCTION

The radial motion of a cylindrical piezoelectric tube and pressure wave propagation inside a chamber are key features of a drop-on-demand inkjet print head. Bogy and Talke[1] found that the basic operating mechanism in an inkjet print head depends on its acoustic characteristics, where the length of the print head and the wave speed of the fluid are key factors to determine these. Acoustic characteristics in a chamber depend on constructive and destructive pressure wave interference. Pressure waves generated by the radial motion of a cylindrical piezoelectric actuator travel toward both ends and they are reflected in- or out-of-phase and interfere with newly generated or residual pressure waves. By adjusting the opening and closing time of a cylindrical piezoelectric actuator, traveling, residual, and newly generated pressure waves interfere with each other constructively or destructively: carefully modulated constructive interference yields an optimum condition for drop formation.

To predict pressure wave generation by the radial motion of a cylindrical piezoelectric actuator, Bugdayci *et al.*[2] developed a simple quasi-static analytic solution and Peelamedu *et al.*[3] compared numerical results with Bugdayci *et al.*'s[2] results and showed good agreement. However, Bugdayci *et al.*'s[2] model is based on a one-dimensional analysis that is only dependent on the radial direction r, and this model does not explain the propagation of pressure waves along the axial direction z, of the cylindrical actuator. Rogers[4] developed a model that accounts for pressure wave propagation in the axial direction but this used a membrane approximation where the radial thickness of the piezoelectric tube is artificially zero.

Meanwhile, Dijksman[5,6] developed an analytic model for pressure wave generation and acoustic interference in a small tubular pump, induced by the radial motion of a cylindrical piezoelectric actuator. The applied voltage signal sets the piezoelectric actuator in motion, which initiates the radial motion of the fluid-filled chamber wall. The square wave driving voltage signal was decomposed into multiple sinusoidal signals and the resultant pressure wave inside the chamber represented by Fourier series terms. However, the radial motion of the inner tube was based on quasi-static thin-shell theory and the radial motion of the piezoelectric actuator was not influenced by the pressure loading inside the chamber. This can lead to error when a thick-walled and/or a compliant inner tube is used with a small but influential pressure loading in a print head. Any pressure within the fluid chamber will resist the radial motion of the chamber wall and thus reduce the magnitude of the axial pressure wave. This effect will be greater if a compliant, polymeric chamber wall, such as a PTFE tube, is used.

The acoustic characteristics of a vibrating cylindrical transducer were investigated by Antohe and Wallace.[7] They used an equivalent circuit model to relate the measured impedance of a cylindrical piezoelectric actuator to predict the intrinsic speed of sound in a fluid. However, ignoring the viscosity of the fluid resulted in a singularity at the acoustic fluid-structure resonance frequency. In addition, ignoring viscosity caused a slightly lower speed of sound because they chose the so-called serial resonance frequency, f_s, to be the same as the mechanical resonance frequency, when the resistance produced by the mechanical losses at the resonant

[a)]Author to whom correspondence should be addressed: Dong-Youn Shin, Manchester Materials Science Center, UMIST, Grosvenor St., Manchester M1 7HS, UK, electronic mail: d.shin@stud.umist.ac.uk, Phone: 44-(0)161-200-3569, Fax: 44-(0)161-200-3586.

mode of the transducer is small enough to be negligible. Although the stiffness of the transducer makes the viscoelastic behavior at the resonant mode negligible, Antohe and Wallace's[7] assumption neglects the influence of viscosity in a fluid on the speed of sound.

In this paper, a simple oscillatory, slightly compressible, fluid flow model is developed with respect to fluid-structure interactions in a cylindrical inkjet actuator. This takes account of the influence of viscosity and pressure loading on the radial motion of the actuator. This model is first used to study the condition of mechanical resonance in a system of known fluid properties. It is then used to analyze the experimental data of Antohe and Wallace[7] and our prediction of fluid sound velocity is compared with their analysis. Finally the model is used to determine the speed of sound in naphthalene and the results compared with literature values.

II. MATHEMATICAL FORMULATION

In this section, dynamic solutions for the cylindrical piezoelectric actuator and fluid will be obtained, which take account of the thickness of the actuator and pressure wave propagation in the axial direction.

A. Radial motion of the piezoelectric tube

It has been pointed out by Luck and Agba[8] that the constitutive equations of piezoelectric materials are handled by many authors without explicitly specifying assumptions of either isothermal or adiabatic condition. Though these conditions can be treated in many cases as equivalent, since an adiabatic process with solids gives little change in temperature, Luck and Agba[8] emphasize that the difference can be critical if a material shows strong pyroelectric/piezoelectric effects because the temperature effects become significant. A piezoelectric actuator in a drop-on-demand inkjet print head can be considered as closer to the isothermal condition than the adiabatic one since the exchange of thermal energy may occur and the thermal gradient across the thickness should be small enough to be negligible in practice. Thus, in this analysis, the piezoelectric transducer is considered in an isothermal condition.

The thickness of the transducer between the electrodes is so small that non-uniformity of the electric field in the transducer is negligible. If there is no shear stress and the electric field is simply given to be $E_r=-V/(r_2-r_1)$ where V, r_2 and r_1 are applied voltage, outer and inner radii of the tube respectively, then the governing constitutive and momentum equations are given by,

$$\sigma_r=c_{33}^E\varepsilon_r+c_{13}^E\varepsilon_\theta+c_{13}^E\varepsilon_z-e_{33}^E E_r, \tag{1}$$

$$\sigma_\theta=c_{13}^E\varepsilon_r+c_{11}^E\varepsilon_\theta+c_{12}^E\varepsilon_z-e_{31}^E E_r, \tag{2}$$

$$\sigma_z=c_{13}^E\varepsilon_r+c_{12}^E\varepsilon_\theta+c_{11}^E\varepsilon_z-e_{31}^E E_r, \tag{3}$$

$$\frac{\partial\sigma_r}{\partial r}+\frac{\sigma_r-\sigma_\theta}{r}=\rho\frac{\partial^2 u}{\partial t^2}. \tag{4}$$

σ_r and σ_θ are the radial and circumferential stresses, u is the radial displacement, t is time, c_{ij}^E, e_{ij}^E and ρ are the elastic stiffness, piezoelectric constants, and density of the PZT actuator at constant electric field respectively. We perform separate analyses with the plane strain assumption and the plane stress assumption. We shall assess the validity of these assumptions later on. If the actuator is made of a composite tube such as a PTFE tube inside the PZT layer, then solutions for all layers must be matched together but in the present paper, only a single layer is considered: the mathematical principles behind the composite tube are very similar to those presented here, except that of course e_{ij}^E and anisotropy vanish in PTFE.

1. Plane strain case

If a plane strain assumption is made, then ε_z is dropped from Eqs. (1) and (2). After putting Eqs. (1) and (2) into Eq. (4) and rearranging it, analytical solutions for the radial displacement and radial stress are obtained as shown in Eqs. (6) and (7), where J, Y and S_1 are Bessel functions of the first and second kinds, and the Lommel function,[9] respectively. For convenience, the time dependence, $e^{i\omega t}$, where i and ω are the imaginary number and angular frequency respectively, is suppressed in all equations.

Rewriting

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\left(\alpha_1-\frac{\beta_1}{r^2}\right)\cdot u+\frac{\gamma_1}{r}=0, \tag{5}$$

where

$$\alpha_1=\frac{\rho\omega^2}{c_{33}^E},\quad \beta_1=\frac{c_{11}^E}{c_{33}^E},\quad \gamma_1=\frac{(e_{31}^E-e_{33}^E)}{c_{33}^E}E_r$$

we obtain

$$u=C_1(z)J_{\sqrt{\beta_1}}(\sqrt{\alpha_1}\cdot r)+C_2(z)Y_{\sqrt{\beta_1}}(\sqrt{\alpha_1}\cdot r) - \frac{\gamma_1}{\sqrt{\alpha_1}}S_1(0,\sqrt{\beta_1},\sqrt{\alpha_1}\cdot r), \tag{6}$$

$$\sigma_r=C_1(z)\begin{pmatrix} -\sqrt{\alpha_1}c_{33}^E J_{\sqrt{\beta_1}+1}(\sqrt{\alpha_1}\cdot r) \\ +(c_{33}^E\sqrt{\beta_1}+c_{13}^E)\dfrac{J_{\sqrt{\beta_1}}(\sqrt{\alpha_1}\cdot r)}{r}\end{pmatrix} + C_2(z)\begin{pmatrix} -\sqrt{\alpha_1}c_{33}^E Y_{\sqrt{\beta_1}+1}(\sqrt{\alpha_1}\cdot r) \\ +(c_{33}^E\sqrt{\beta_1}+c_{13}^E)\dfrac{Y_{\sqrt{\beta_1}}(\sqrt{\alpha_1}\cdot r)}{r}\end{pmatrix} + \gamma_1\begin{pmatrix} (c_{33}^E\sqrt{\beta_1}-c_{13}^E)\dfrac{S_1(0,\sqrt{\beta_1},\sqrt{\alpha_1}\cdot r)}{\sqrt{\alpha_1}\cdot r} \\ -c_{33}^E(\sqrt{\beta_1}-1)S_1(-1,\sqrt{\beta_1}-1,\sqrt{\alpha_1}\cdot r)\end{pmatrix} - e_{33}^E E_r. \tag{7}$$

It is noteworthy that unknown z dependent functions, $C_1(z)$ and $C_2(z)$, are involved in Eqs. (6) and (7). The origin of these functions comes from the second order ordinary differential equation with respect to r in Eq. (5). If the main focus is given to the radial motion of the tube, then the longitudinal motion of the tube and the shear stress, which may be much smaller than the radial stress, can be ignored, and the solu-

tion will take the form of Eqs. (6) and (7). Thus, $C_1(z)$ and $C_2(z)$ take account of the wave propagation along the axial direction. These functions will be determined by the pressure boundary conditions on the tube.

2. Plane stress case

If a plane stress assumption is employed, then σ_z is assumed to be 0 and the following differential equation is obtained. The radial displacement and stress solutions take the same forms of Eqs. (6) and (7) except for the definitions of α_1, β_1 and γ_1.

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\left(\alpha_1-\frac{\beta_1}{r^2}\right)\cdot u+\frac{\gamma_1}{r}=0, \tag{8}$$

where

$$a=c_{33}^E-\frac{c_{13}^{E2}}{c_{11}^E}, \qquad b=c_{13}^E-\frac{c_{12}^E c_{13}^E}{c_{11}^E},$$

$$c=\left(\frac{c_{13}^E e_{31}^E}{c_{11}^E}-e_{33}^E\right)E_r,$$

$$d=c_{11}^E-\frac{c_{12}^{E2}}{c_{11}^E}, \quad e=\left(\frac{c_{12}^E e_{31}^E}{c_{11}^E}-e_{31}^E\right)E_r,$$

$$\alpha_1=\frac{\rho\omega^2}{a}, \quad \beta_1=\frac{d}{a}, \quad \gamma_1=\frac{c-e}{a}.$$

B. Analysis for the oscillatory slightly compressible fluid motion of the actuator

If the fluid flow inside a thick-walled elastic tube is assumed to be axisymmetric and laminar flow, then the θ momentum Navier-Stokes equation becomes trivial and only mass continuity and r and z momentum equations remain. Only small density changes are allowed about the mean fluid density, ρ_f, and the continuity equation is linearized in the velocity and the small density change. This should be a good approximation for a liquid. In addition, pressure inside a small tube, which is generated by the radial motion of the piezoelectric tube, can be assumed to be independent of the r direction because of the high aspect ratio of the PZT tube. Moreover, because the characteristic wave speed in a tube is fast and its wavelength is fairly long compared to the dimensions of the tube, the second derivative term, $\partial^2 V_z/\partial z^2$, where V_z is the axial velocity, can be discarded in the z momentum Navier-Stokes equation. With the assumptions made above, the integration of the mass continuity equation over the cross-section yields Eq. (9):

$$\frac{1}{\rho_f}\frac{\partial \rho_f'}{\partial t}+\frac{1}{r}\frac{\partial}{\partial r}(rV_r)+\frac{\partial V_z}{\partial z}=0,$$

where ρ_f' denotes actual fluid density $|\rho_f'-\rho_f|\ll\rho_f$ and

$$\frac{1}{\rho_f}\frac{\partial \rho_f'}{\partial t}=\frac{1}{\rho_f c_f^2}\frac{\partial \overline{P}(z,t)}{\partial t}=\frac{i\omega}{\rho_f c_f^2}P(z)$$

$$\frac{i\omega\pi r_1^2}{\rho_f c_f^2}P(z)+2\pi r_1 V_r|_{r_1}+\frac{\partial}{\partial z}\left(\int_0^{r_1}2\pi r V_z dr\right)=0, \tag{9}$$

where c_f and $P(z)$ are the intrinsic speed of sound in a fluid and the z dependent pressure function.

By integrating the continuity equation over the cross-section, the detailed calculation of the r momentum equation is not necessary. At the interface between the fluid and the structure, the differential of the radial motion of the piezoelectric tube with respect to time becomes $V_r|_{r_1}$, where V_r is the radial velocity. The stress boundary conditions are given by zero at the exterior of the tube, $r=r_2$, and the fluid pressure at the inner surface, $r=r_1$. Or more formally,

$$\sigma_r|_{r_2}=c_{33}^E\left.\frac{\partial u}{\partial r}\right|_{r_2}+c_{13}^E\left.\frac{u}{r}\right|_{r_2}-e_{33}^E E_r=0, \tag{10}$$

$$\sigma_r|_{r_1}=c_{33}^E\left.\frac{\partial u}{\partial r}\right|_{r_1}+c_{13}^E\left.\frac{u}{r}\right|_{r_1}-e_{33}^E E_r=-P(z). \tag{11}$$

From which, the radial displacement and velocity functions become:

$$u|_{r_1}=(j_1+k_1P(z)), \tag{12}$$

$$V_r|_{r_1}=i\omega(j_1+k_1P(z)). \tag{13}$$

The coefficients, j_1 and k_1, are defined in the Appendix. At this point, the unknown z dependent functions, $C_1(z)$ and $C_2(z)$, are determined in terms of the propagating pressure wave amplitude. After substituting Eq. (13) into Eq. (9) and rearranging it, Eq. (14) is obtained.

$$\frac{\partial}{\partial z}\left(\int_0^{r_1}2\pi r V_z dr\right)+\left(2\pi r_1 i\omega k_1+\frac{i\omega\pi r_1^2}{\rho_f c_f^2}\right)P(z)$$

$$+2\pi r_1 i\omega j_1=0. \tag{14}$$

If V_z is composed of the product of r and z dependent functions, $f(r)$, $g(z)$, and a time function, $e^{i\omega t}$, and we search for solutions where $g(z)$ takes the form of an exponential function, then the linearized z momentum Navier-Stokes equation becomes as shown in Eq. (15) and its solution, which has a no-slip condition at $r=r_1$, is given in Eq. (16):

$$\frac{\partial^2 f(r)}{\partial r^2}+\frac{1}{r}\frac{\partial f(r)}{\partial r}+\left(-\frac{i\omega}{\nu}\right)f(r)-\frac{C^*}{\rho_f\nu}=0, \tag{15}$$

$$f(r)=\frac{-C^*}{\rho_f i\omega}\left(1-\frac{J_0\left(\sqrt{-\frac{i\omega}{\nu}}r\right)}{J_0\left(\sqrt{-\frac{i\omega}{\nu}}r_1\right)}\right), \tag{16}$$

$$V_z=\frac{-1}{\rho_f i\omega}\left(1-\frac{J_0\left(\sqrt{-\frac{i\omega}{\nu}}r\right)}{J_0\left(\sqrt{-\frac{i\omega}{\nu}}r_1\right)}\right)\frac{\partial P(z)}{\partial z}, \tag{17}$$

where C^* is a constant given by $(\partial P(z)/\partial z)/g(z)$ and ν is the kinematic viscosity. After integrating Eq. (17) over the cross-section we obtain Eq. (18) and substituting this into Eq. (14), it yields Eq. (19).

TABLE I. Properties of ethylene glycol and isopropyl alcohol (Refs. 7, 13, 14).

	Ethylene glycol	Isopropyl alcohol
Density [$\mathrm{kg\cdot m^{-3}}$], ρ_f	1113	785.5
Viscosity [$\mathrm{Pa\cdot s}$], μ	2×10^{-2}	2.43×10^{-3}
Speed of sound [$\mathrm{m\cdot s^{-1}}$], c_f	1650	1170

$$\frac{\partial}{\partial z}\left(\int_0^{r_1} 2\pi r V_z\, dr\right) = \frac{-\pi}{\rho_f i\omega}\left(r_1^2 - \frac{2r_1\cdot J_1\left(\sqrt{-\frac{i\omega}{\nu}}r_1\right)}{\sqrt{-\frac{i\omega}{\nu}}\cdot J_0\left(\sqrt{-\frac{i\omega}{\nu}}r_1\right)}\right)\frac{\partial^2 P(z)}{\partial z^2}, \tag{18}$$

$$\frac{-\pi}{\rho_f i\omega}\left(r_1^2 - \frac{2r_1\cdot J_1\left(\sqrt{-\frac{i\omega}{\nu}}r_1\right)}{\sqrt{-\frac{i\omega}{\nu}}\cdot J_0\left(\sqrt{-\frac{i\omega}{\nu}}r_1\right)}\right)\frac{\partial^2 P(z)}{\partial z^2} + \left(2\pi r_1 i\omega k_1 + \frac{i\omega\pi r_1^2}{\rho_f c_f^2}\right)P(z) + 2\pi r_1 i\omega j_1 = 0. \tag{19}$$

Equation (19) will be solved for the pressure function, $P(z)$, with given boundary conditions at both ends of the tube. Since most inkjet print heads operate at frequencies lower than the first structural resonance frequency, most of the energy can be assumed to be reflected from open ends. Though it is quite complicated to obtain an exact boundary condition for a flanged or unflanged open end, an open-end correction factor for the finite length of the tube given by Rossing and Fletcher,[10] and Pierce[11] is simply employed as follows:

$$l_e = l + \Delta l_{left\_end} + \Delta l_{right\_end} = l + 2\times 0.61\times r_1, \tag{20}$$

where l and l_e are the actual and effective length of the tube.

III. RESULTS AND DISCUSSION

The model is first used to simulate the behavior of a piezoelectric tube of PZT 5A with two different fluids, isopropyl alcohol and ethylene glycol, for the purpose of validation and comparison with Antohe and Wallace's[7] work. It is assumed that the piezoelectric tube is immersed in a fluid without any air bubble entrapment and has the same dimensions as the tube used by Antohe and Wallace,[7] with $r_1 = 0.41$ mm, $r_2 = 0.65$ mm and $l = 15.33$ mm. The thickness of the electrodes is assumed to be small enough to be ignored. The properties of the fluids are listed in Table I.

Figure 1 shows the real and imaginary pressure coefficients with plane strain and stress assumptions (all pressure coefficients being computed at the center of the PZT tube). The resonance frequency measured by Antohe and Wallace,[7] when a PZT tube is submerged in ethylene glycol, is 45.08 kHz and the computed resonance frequency with plane strain and stress assumptions are 45.8 kHz and 45.5 kHz respectively. The resonance frequencies by Antohe and Wallace,[7] plane strain and plane stress assumptions, when a PZT tube is submerged in isopropyl alcohol, are 34.47 kHz, 35.1 kHz and 35 kHz, respectively.

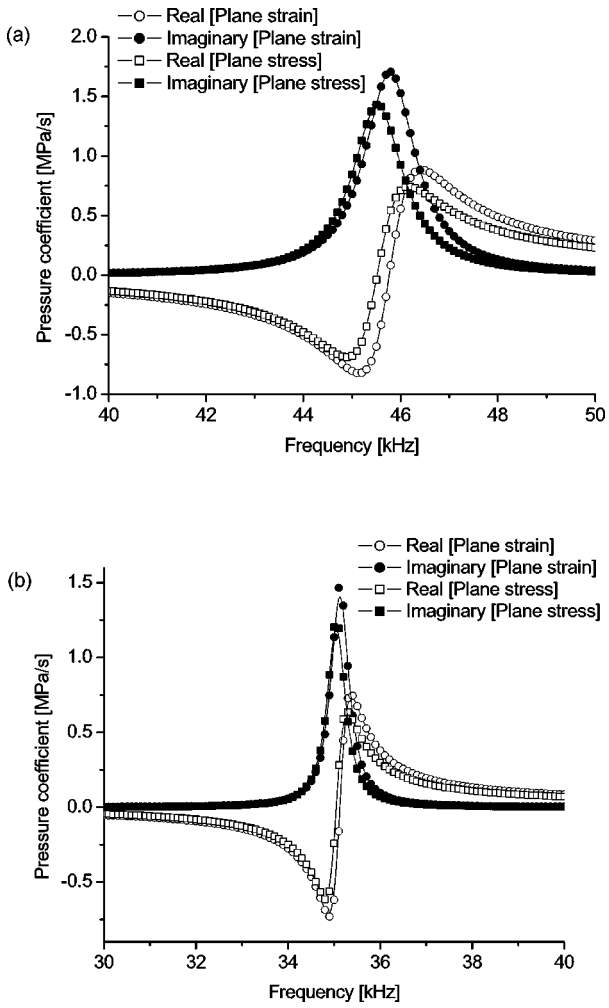


FIG. 1. Comparison of plane strain and plane stress assumptions submerged in (a) Ethylene glycol. (b) Isopropyl alcohol.

If the radial displacement and velocity of the PZT transducer are represented by Eqs. (21) and (22) and Eq. (22) is put into the continuity equation (14) as shown in Eq. (23), then the wave speed in a compliant conduit is defined to be the term in the first parentheses and written in Eq. (24):

$$u = j_1 + k_1\cdot P, \tag{21}$$

$$V_r = \frac{\partial j_1}{\partial t} + k_1\cdot\frac{\partial P}{\partial t}, \tag{22}$$

$$\left(\frac{1}{c_f^2} + \frac{2\rho_f k_1}{r_1}\right)\frac{\partial P}{\partial t} + \frac{2\rho_f}{r_1}\frac{\partial j_1}{\partial t} + \frac{2\rho_f}{r_1^2}\int_0^{r_1} r\frac{\partial V_z}{\partial z}dr = 0, \tag{23}$$

$$a_f = \sqrt{\frac{c_f^2 r_1}{r_1 + 2\rho_f k_1 c_f^2}}. \tag{24}$$

This wave speed derivation, as stated in the work of Dijksman,[5] is preferred by authors to Eq. 2.35 of Wylie and Streeter.[12] This derivation is straightforward and suitable for the estimation of the wave speed in a composite tube, while

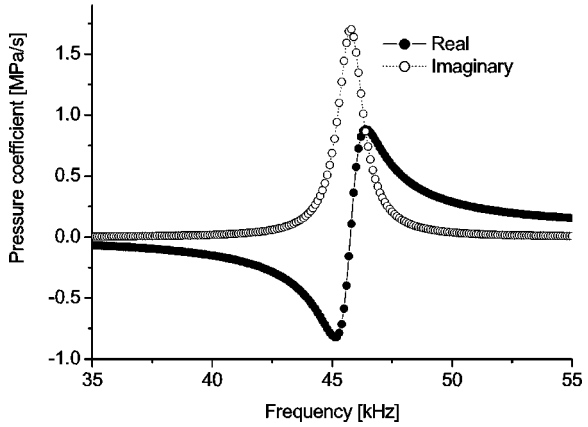


FIG. 2. Pressure coefficient response, generated by the radial motion of the PZT tube in ethylene glycol, against frequency at 5 V.

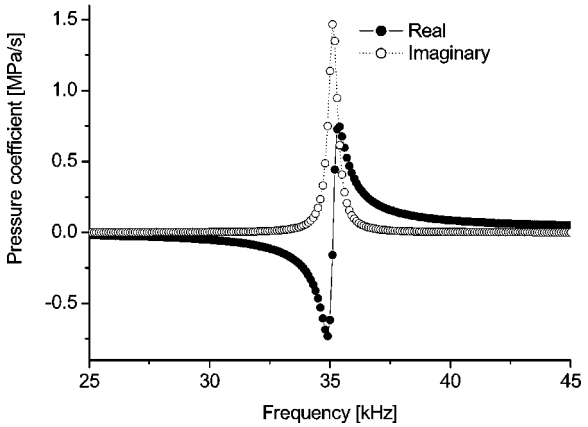


FIG. 3. Pressure coefficient response, generated by the radial motion of the PZT tube in isopropyl alcohol, against frequency at 5 V.

the equations of Wylie and Streeter[12] may not be suitable for an anisotropic conduit.

According to the given boundary conditions at the outer surface and ends of a PZT tube, if a PZT tube is submerged in ethylene glycol, simple calculations are done to investigate the influence of plane strain and plane stress assumptions to find out the resonance frequency.

Wave speeds with the plane strain and plane stress assumptions are approximately $1469.6\,\mathrm{m\cdot s^{-1}}$ and $1461.6\,\mathrm{m\cdot s^{-1}}$, respectively, and resonant frequencies are 46.4 kHz and 46.2 kHz, respectively. The wave speed is slightly affected by frequency but this has very little effect and is hence negligible for the above calculations. In terms of acoustic resonance frequency, these two assumptions do not produce any significant differences.

Radial and axial strain, ε_r and ε_z, are compared when a plane stress assumption is adopted under static and traction free conditions. At the inner surface of a PZT tube, the absolute ratio of ε_z to ε_r is 0.28. The ratio of these two strains is smaller than unity and so the adoption of a plane strain assumption seems a reasonable simplification. The proper assumption in the case of a tube with finite length cannot be made easily and it is not known to the authors if there is any specific aspect ratio of radial thickness to length of a tube, which can help to decide which assumption is valid.

However, with the facts that the ratio of ε_z to ε_r at the inner surface of a PZT tube, where a liquid interact with a PZT tube, is still smaller than unity, the ratio of the radial thickness to the length of a PZT tube is very low, and the resonant frequencies are not significantly affected with different assumptions, a plane strain assumption in the z direction is chosen for simplicity.

In order to study the acoustic response of the system, the range of frequencies from 10 kHz to 80 kHz are swept computationally with an increment of 100 Hz at an amplitude of 5 V. The computation results as a function of frequency of the real and imaginary pressure coefficients caused by the radial motion of the PZT 5A tube are shown in Figs. 2 and 3.

The neglect of the viscoelastic behavior of a PZT actuator, which leads to a somewhat high but still finite pressure amplitude due to viscous dissipation of the fluid in its resonance mode, does not affect the detection of the fluid-structure resonance frequency. Serial and parallel resonance frequencies, f_s and f_p, respectively, are still evident as the minimum and maximum of the real part of the pressure response. However, the incorporation of viscosity shows its effect due to the appearance of an imaginary component of pressure. The peak frequency of the imaginary coefficient of the pressure corresponds to the peak phase angle frequency. The analogy between predicted real/imaginary coefficients of pressure and measured impedance/phase angle versus frequency is shown in Fig. 4.

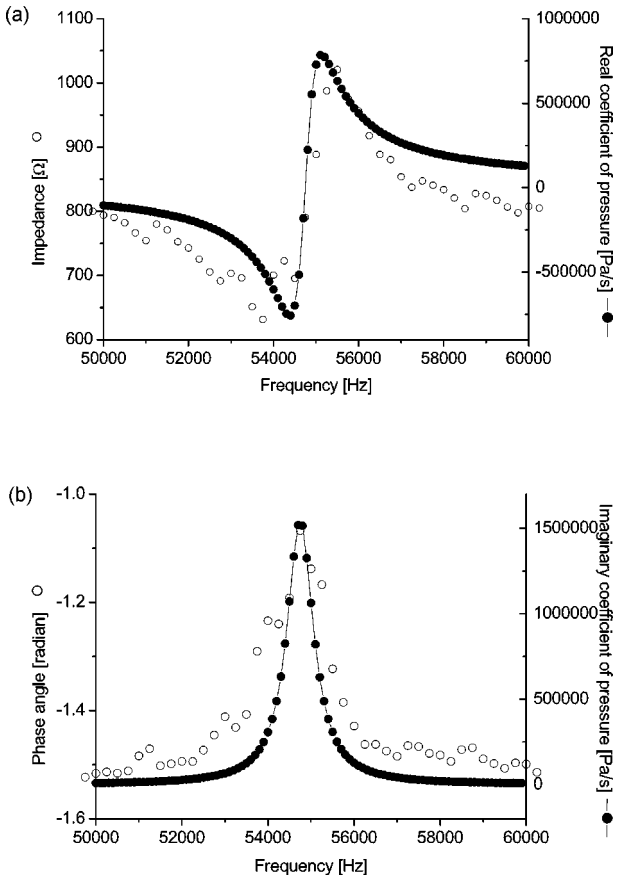


FIG. 4. Analogy between real/imaginary coefficients of pressure and measured impedance/phase angle against frequency, with 40 volume % alumina suspension at 120 °C. (a) The real coefficient of pressure and impedance. (b) The imaginary coefficient of pressure and phase angle.

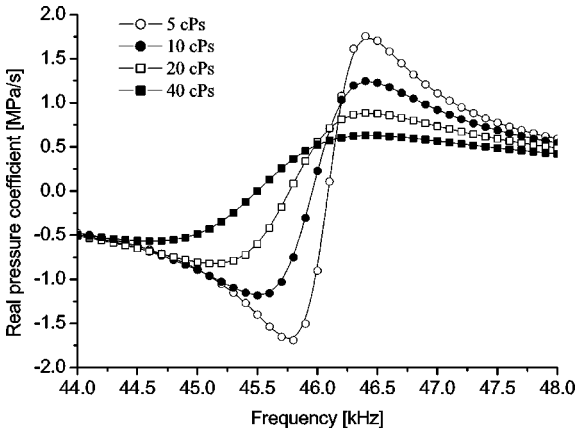


FIG. 5. Influence of viscosity on the acoustic resonance frequency.

The gap between serial and parallel resonance frequencies tends to broaden as viscosity increases. To make the influence of viscosity on the acoustic resonance frequency clear, a fictitious fluid having the same properties as ethylene glycol except for viscosity is introduced. Figure 5 shows the resonance frequencies become smaller and the gap between serial and parallel resonance frequencies becomes wider as the viscosity of a liquid increases, though the parallel resonance frequencies are almost fixed. This implies that the neglect of viscosity in a fluid can cause an error in terms of the determination of the speed of sound from impedance measurements on a PZT actuator. When the viscosity term is neglected by assuming an inviscid fluid, then the imaginary coefficient does not appear and serial and parallel resonance frequencies become the same since a singularity appears in the resonant mode. The actual resonance occurs when real and imaginary coefficients of pressure are both considered and this implies that the choice of the serial resonance frequency for the determination of the speed of sound in a fluid is error prone. Instead, the frequency corresponding to the peak phase angle, which is generally located between serial and parallel resonance frequencies, should be selected for the computation of the speed of sound using the impedance measurement method because the absolute magnitude of the pressure coefficients, $\sqrt{real(P)^2+imag(P)^2}$, becomes maximal near there.

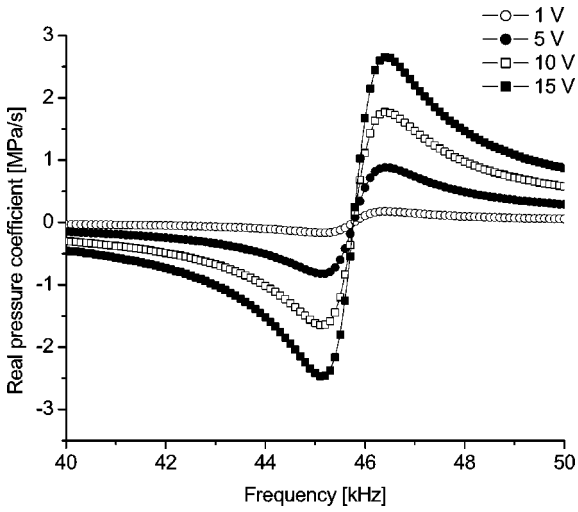


FIG. 6. Influence of applied voltage on the acoustic resonance frequency.

TABLE II. Comparison of acoustic resonance frequencies.

Measurement	Acoustic resonance frequency [kHz]	
	Ethylene glycol	Isopropyl alcohol
Measured by Antohe and Wallace (Ref. 7)	45.08	34.47
Computed	45.78	35.13

Figure 6 shows computed resonance frequencies with different voltage when a PZT tube is submerged in ethylene glycol. In a PZT tube, a standing wave forms and it is virtually independent of applied voltage except for the amplitude of the pressure wave and hence resonance frequency is not affected.

The acoustic fluid-structure resonance frequencies measured by Antohe and Wallace,[7] and the computed resonance frequencies using the current model are listed in Table II for a pair of fluids. Errors of computed acoustic resonance frequencies with the developed model to measured frequencies are $+1.55\%$ and $+1.91\%$, respectively. This shows good agreement.

So far, it has been assumed that the intrinsic speed of sound, c_f, in a fluid is given and the acoustic fluid-structure resonance frequency, f_r, which corresponds to the peak phase angle frequency, is computed and compared to experimentally measured and expected frequencies. However, in practice, the intrinsic speed of sound is not known but the acoustic fluid-structure resonance frequency can be easily measured with an impedance analyzer. After substituting the measured resonance frequency into the angular frequency, $\omega=2\pi\cdot f_r$, of the current model, a speed of sound sweep was performed. It is expected that at a certain speed of sound, the absolute magnitude of the pressure coefficients, $\sqrt{real(P)^2+imag(P)^2}$, attains a maximum peak. The predicted intrinsic speed of sound is chosen where the maximum peak appears. The sweep range for the guessed speed of sound is 1000 $\mathrm{m\cdot s^{-1}}$ to 2000 $\mathrm{m\cdot s^{-1}}$ with 5 $\mathrm{m\cdot s^{-1}}$ increment and results are shown in Fig. 7. Then the predicted speeds of sound in ethylene glycol and isopropyl alcohol

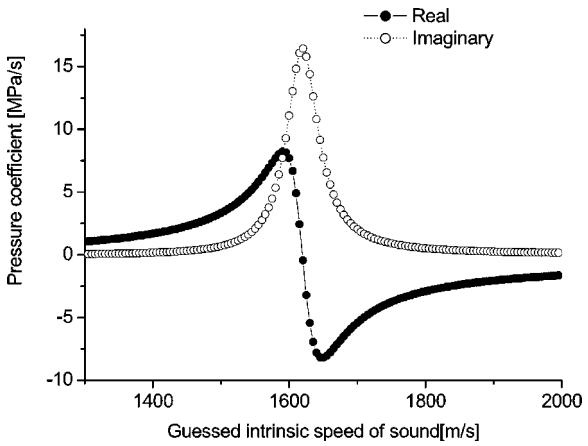


FIG. 7. Determination of the speed of sound with measured resonance frequencies in ethylene glycol at 5 V.

TABLE III. Relative errors with respect to measured and computed speed of sound.

	Error [%]		
	Ethylene glycol	Isopropyl alcohol	Naphthalene
Antohe and Wallace's (Ref. 7) model	−3.1	−2.7	−1.56
Current model	−1.81	−2.14	−1.06

correspond to $1620\,\mathrm{m\cdot s^{-1}}$ and $1145\,\mathrm{m\cdot s^{-1}}$, respectively. This compares favorably with the literature values shown in Table I.[13,14]

To investigate the influence of temperature on the measurement of the speed of sound in a fluid, the impedance of the transducer in naphthalene[15] at 100 °C, of which the speed of sound is $1250\,\mathrm{m\cdot s^{-1}}$ at the stated temperature, is measured. The detailed procedure for the impedance measurement is the same as that of Antohe and Wallace.[7] The physical material constants of a piezoelectric transducer change with temperature but the result presented in naphthalene at 100 °C shows the speed of sound in naphthalene has only −1.06% difference from the value reported in the literature. Relative errors with respect to literature-reported speeds of sound are listed in Table III.

IV. CONCLUSION

Using experimentally measured and theoretically predicted acoustic fluid-structure resonance frequencies, an oscillatory, slightly compressible, flow in a thick-walled piezoelectric tube immersed in a fluid is evaluated. It is shown that the inclusion of the fluid's viscosity produces the expected behavior of real and imaginary coefficients of the pressure, corresponding to measured impedance/phase angle curves, and as viscosity increases, the difference between serial and parallel resonance frequencies increases. Due to the analogies of real coefficient of pressure-impedance and imaginary coefficient of pressure-phase angle, it is proposed to select the peak phase angle frequency where the overall magnitude of real and imaginary coefficients of pressure reaches its maximum. Though physical properties of the piezoelectric transducer change with temperature, it is shown that the quantitative result used to determine the speed of sound in a fluid is not much affected even above 100 °C. The model provides good agreement with literature values (within 2%) when it is used to determine the speed of sound from acoustic resonance.

ACKNOWLEDGMENTS

We acknowledge the support of the EPSRC through Project No. GR/N16969. D-Y. Shin would like to specially thank Rotary International for a postgraduate scholarship.

APPENDIX: COEFFICIENTS

J, Y and S_1 stand for Bessel functions of the first and second kinds, and the Lommel function. The following constants are defined:

$$a_1 = -\sqrt{\alpha_1}c_{33}^E J_{\sqrt{\beta_1}+1}(\sqrt{\alpha_1}\cdot r_2) + (c_{33}^E\sqrt{\beta_1}+c_{13}^E)\frac{J_{\sqrt{\beta_1}}(\sqrt{\alpha_1}\cdot r_2)}{r_2},$$

$$b_1 = -\sqrt{\alpha_1}c_{33}^E Y_{\sqrt{\beta_1}+1}(\sqrt{\alpha_1}\cdot r_2) + (c_{33}^E\sqrt{\beta_1}+c_{13}^E)\frac{Y_{\sqrt{\beta_1}}(\sqrt{\alpha_1}\cdot r_2)}{r_2},$$

$$c_1 = \gamma_1\begin{pmatrix} (c_{33}^E\sqrt{\beta_1}-c_{13}^E)\dfrac{S_1(0,\sqrt{\beta_1},\sqrt{\alpha_1}\cdot r_2)}{\sqrt{\alpha_1}\cdot r_2} \\ -c_{33}^E(\sqrt{\beta_1}-1)S_1(-1,\sqrt{\beta_1}-1,\sqrt{\alpha_1}\cdot r_2)\end{pmatrix} - e_{33}^E E_r,$$

$$d_1 = -\sqrt{\alpha_1}c_{33}^E J_{\sqrt{\beta_1}+1}(\sqrt{\alpha_1}\cdot r_1) + (c_{33}^E\sqrt{\beta_1}+c_{13}^E)\frac{J_{\sqrt{\beta_1}}(\sqrt{\alpha_1}\cdot r_1)}{r_1},$$

$$e_1 = -\sqrt{\alpha_1}c_{33}^E Y_{\sqrt{\beta_1}+1}(\sqrt{\alpha_1}\cdot r_1) + (c_{33}^E\sqrt{\beta_1}+c_{13}^E)\frac{Y_{\sqrt{\beta_1}}(\sqrt{\alpha_1}\cdot r_1)}{r_1},$$

$$f_1 = \gamma_1\begin{pmatrix} (c_{33}^E\sqrt{\beta_1}-c_{13}^E)\dfrac{S_1(0,\sqrt{\beta_1},\sqrt{\alpha_1}\cdot r_1)}{\sqrt{\alpha_1}\cdot r_1} \\ -c_{33}^E(\sqrt{\beta_1}-1)S_1(-1,\sqrt{\beta_1}-1,\sqrt{\alpha_1}\cdot r_1)\end{pmatrix} - e_{33}^E E_r,$$

$$g_1 = J_{\sqrt{\beta_1}}(\sqrt{\alpha_1}\cdot r_1), \quad h_1 = Y_{\sqrt{\beta_1}}(\sqrt{\alpha_1}\cdot r_1),$$

$$i_1 = -\frac{\gamma_1}{\sqrt{\alpha_1}}S_1(0,\sqrt{\beta_1},\sqrt{\alpha_1}\cdot r_1),$$

$$j_1 = \frac{(b_1f_1-c_1e_1)g_1+(d_1c_1-a_1f_1)h_1}{a_1e_1-d_1b_1}+i_1,$$

$$k_1 = \frac{b_1g_1-a_1h_1}{a_1e_1-d_1b_1}.$$

[1] D.B. Bogy and F.E. Talke, "Experimental and theoretical study of wave propagation phenomena in drop-on-demand ink jet devices," IBM J. Res. Dev. **28**, 314–321 (1984).

[2] N. Bugdayci, D.B. Bogy, and F.E. Talke, "Axisymmetric motion of radially polarized piezoelectric cylinder used in ink jet printing," IBM J. Res. Dev. **27**, 171–180 (1983).

[3] S.M. Peelamedu, C.B. Kosaraju, R.V. Dukkipati, and N.G. Naganathan, "Numerical Approach for Axisymmetric piezoceramic geometries towards fluid control applications," P. I. Mech. Eng. I-J.Sys. **214**, 87–97 (2000).

[4] P.H. Rogers, "Mathematical model for a free-flooded piezoelectric cylinder transducer," J. Acoust. Soc. Am. **80**, 13–18 (1986).

[5] J.F. Dijksman, "Hydrodynamics of small tubular pumps," J. Fluid Mech. **139**, 173–191 (1984).

[6] J.F. Dijksman, "Hydro-acoustics of piezoelectrically driven ink-jet print heads," Flow, Turbul. Combust. **61**, 211–237 (1999).

[7] B.V. Antohe and D.B. Wallace, "The determination of the speed of sound in liquids using acoustic resonance in piezoelectric tubes," Meas. Sci. Technol. **10**, 994–998 (1999).

[8] R. Luck and E.I. Agba, ''On the design of piezoelectric sensors and actuators,'' ISA Trans. **37**, 65–72 (1998).
[9] Y. Luke, *The Special Functions and their Approximations, Vol. I* (Academic, New York, 1969).
[10] T.D. Rossing and N.H. Fletcher, *The Physics of Musical Instruments*, 2nd ed. (Springer, New York, 1998), p. 200.
[11] A.D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (The Acoustical Society of America, New York, 1989), p. 348.
[12] E.B. Wylie and V.L. Streeter, *Fluid Transients in Systems* (Prentice-Hall, New Jersey, 1993).
[13] R.C. Weast, S.M. Selby, and C.D. Hodgman, *Handbook of Chemistry and Physics*, 53rd ed. (Chemical Rubber Co., Cleveland, OH, 1972).
[14] W.P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics* (Van Nostrand, New York, 1950).
[15] S.J. Pugh and R.F. Lambert, *Fluid Transients in Pipes and Tunnel: Speed of Propagation of Pressure Waves* (Engineering Sciences Data Unit (ESDU) Data item 83046, 2001).

Fast Fourier transform and singular value decomposition formulations for patch nearfield acoustical holography

Earl G. Williams[a] and Brian H. Houston
Code 7130, Naval Research Laboratory, Washington, DC 20375

Peter C. Herdic
SFA, Inc., Landover, Maryland 20785

(Received 26 December 2002; revised 31 May 2003; accepted 7 July 2003)

Nearfield acoustical holography (NAH) requires the measurement of the pressure field over a complete surface in order to recover the normal velocity on a nearby concentric surface, the latter generally coincident with a vibrator. Patch NAH provides a major simplification by eliminating the need for complete surface pressure scans—only a small area needs to be scanned to determine the normal velocity on the corresponding (small area) concentric patch on the vibrator. The theory of patch NAH is based on (1) an analytic continuation of the patch pressure which provides a spatially tapered aperture extension of the field and (2) a decomposition of the transfer function (pressure to velocity and/or pressure to pressure) between the two surfaces using the singular value decomposition (SVD) for general shapes and the fast Fourier transform (FFT) for planar surfaces. Inversion of the transfer function is stabilized using Tikhonov regularization and the Morozov discrepancy principle. Experimental results show that root mean square errors of the normal velocity reconstruction for a point-driven vibrator over 200–2700 Hz average less than 20% for two small, concentric patch surfaces 0.4 cm apart. Reconstruction of the active normal acoustic intensity was also successful, with less than 30% error over the frequency band. [DOI: 10.1121/1.1603767]

PACS numbers: 43.20.Ye, 43.40.At, 43.60.Pt [MO]

I. INTRODUCTION

Nearfield acoustical holography (NAH) provides a reconstruction of the normal velocity on a surface from a measurement of the radiated field on a near concentric surface. This reconstruction is implemented by solving an inverse problem with its roots in the Rayleigh integral for planar geometries, and the Helmholtz integral equation for more complex geometries. The Rayleigh and Helmholtz integrals represent forward problems, requiring a knowledge of the surface fields over a complete boundary to compute the radiated pressure anywhere on the near surface. In the same way the inverse problem requires the pressure field on the complete near surface (infinite surface for the planar case) to reconstruct the normal velocity at any point on the source surface. For example to reconstruct the normal velocity on the surface of a sphere NAH requires that the pressure field be measured over the complete spherical surface located exterior to the vibrator. In the strict sense NAH theory does not apply when the pressure is known only over a partial surface. Thus the pressure on one of the polar caps is generally not sufficient for an accurate reconstruction of the normal velocity on the corresponding vibrating polar cap. The objective of this paper is to overcome this restriction through a modification of NAH theory which provides accurate reconstructions of the normal velocity from patch pressure data when the patch is considerably smaller than the vibrating surface. In this effort we depart from rigorous theory, realizing that the current objective is to determine a minimum error approximation of the normal velocity field conformal to the patch pressure data.

There are three approaches in current NAH theory to solving the inverse problem: (1) The Fourier transform implemented with the discrete Fourier transform (DFT), used in simple geometries[1–4] such as planes and cylinders, including implementations for noncoherent sources,[5] (2) the singular value decomposition (SVD) in conjunction with inverse boundary elements (IBEM) generally used for more complex geometries,[6–15] and (3) NAH approaches based upon spherical wave expansions which do not use IBEM also deal with complicated geometries.[16,17] We will concentrate on the first two of these approaches in this paper.

IBEM depends upon computation a priori of the transfer function between the measured pressure and the desired reconstruction quantity (usually normal velocity or surface pressure). This transfer function is computed through theoretical means when the geometry lends itself to construction of the analytic form of the Green function, as is the case, for example, for a cylindrical cavity;[18] or through the use of boundary element methods (BEM) when no analytic form exists for the measurement geometry.[10,19] Some IBEM–SVD approaches have been used for planar problems.[20–22] These planar applications use a truncated aperture (large patch) which is the size of the actual vibrator which differs from our approach which deals with much smaller patches. An early approach[23] did use a measurement patch much smaller than the vibrator, however, the formulation was based on an axisymmetric structure and required SVD inversion of the Green function operator relating the pressure on the measure-

[a]Author to whom correspondence should be addressed; electronic mail: williams@pa.nrl.navy.mil

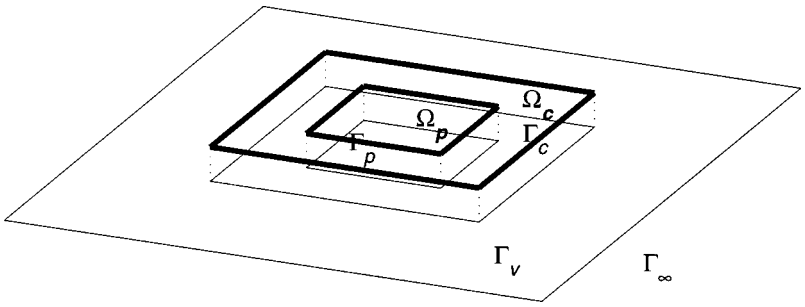


FIG. 1. Definitions of the domains for the source plane and the measurement plane. Ω_p defines the patch region (the measurement hologram) and is surrounded by the "picture frame" Ω_c of constant width used for the analytic continuation of the hologram. Similarly, Γ_p and Γ_c are defined in the plane of the vibrator surface. The vibrator surface is indicated by $\Gamma_v \in \Gamma_\infty$.

ment patch to the normal velocity *at all points* on the structure surface. An approach which avoids the FFT and uses patches smaller than the vibrator was developed by Hald for planar geometry.[24,25] Another approach[13] reduced the number of measurement points from a set extending past the vibrator using the effective independence method to determine a reduced set of optimal measurement locations, although not necessarily on a smaller patch. Recently a patch approach based on an iteration procedure appeared[26] using FFT based NAH. We build upon this approach here extending it to SVD based NAH as well as improving the accuracy of the reconstructions for FFT and SVD implementations by adding regularization theory to the inversions.

We begin with a discussion of the analytic continuation approach which is used to avoid serious problems at the edges of the surface velocity reconstructions. Discussed in a separate paper,[27] but reviewed briefly in Sec. III, this continuation approach forms the cornerstone of the method here and is applied not only to the FFT version of NAH but also the SVD version, allowing development of patch NAH for general geometries. In Sec. III C we provide experimental results of the patch approach studying separately the FFT and the SVD approaches to NAH. The results compare the reconstruction of the normal surface velocity from a patch compared to the exact result and determine the errors over a wide band of frequencies. Also presented is the accuracy of normal intensity reconstructions on the vibrator surface, again using patch measurements and the FFT and SVD approaches.

II. PATCH NAH USING FOURIER ACOUSTICS AND THE SVD

Before discussion of the new patch NAH algorithm, we review the basic theory of Fourier transform based NAH and regularization theory used for inversion, in order to illuminate the finite aperture replication (AR) and edge discontinuity effects which are the major nemeses for development of patch NAH. In Sec. II C below we discuss the SVD approach to inversion of Eq. (1) and compare it with the Fourier transform based approach. The SVD method is extremely important because it is not restricted to separable geometries.

A. Review of Fourier transform based NAH

We solve the following inverse problem: determine the normal surface velocity on a plane Γ, normally containing a vibrating planar source, given the radiated acoustic pressure in a parallel plane Ω located a small distance away. The normal projection of Ω onto the source plane yields Γ. As shown in Fig. 1 we denote the infinite source plane Γ_∞ and the vibrator surface as Γ_v. This problem is central to nearfield acoustical holography and requires the inversion of the integral equation relating the pressure and normal velocity given in the case of planar geometry by Rayleigh's first integral formula

$$p(\mathbf{x}) = \frac{-i\rho ck}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dot{w}(\mathbf{y}) \frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|} d\Gamma(\mathbf{y}), \tag{1}$$

where $p(\mathbf{x})$, $\mathbf{x}=(x_1,x_2,x_3)$ with $x_3=d\geqslant 0$, is the pressure measured on an infinite plane sheet $\mathbf{x}\in\Omega_\infty$ a distance d from the source plane; $\dot{w}(\mathbf{y})$, $\mathbf{y}=(y_1,y_2,y_3)$ with $y_3=0$, is the unknown normal velocity in the source plane and the integration is over the infinite plane $\mathbf{y}\in\Gamma_\infty$ containing the source Γ_v. The harmonic time dependence $e^{-i\omega t}$ is suppressed throughout this paper, k is the acoustic wave number and ρc is the impedance of the medium. When the pressure $p(\mathbf{x})$ is measured close to the vibrator and is recorded over an area larger than the vibrating surface Γ_v, NAH is extremely successful at inverting Eq. (1) accurately.[4] If on the other hand the pressure is known only over a small patch, $\mathbf{x}\in\Omega_p$ ($\Omega_p \in \Omega_\infty$), above the vibrating surface, NAH is not usually successful in generating accurate reconstructions. See Fig. 1. We investigate in this paper methods which are robust for the inversion of Eq. (1) when $p(\mathbf{x})$ is known only in Ω_p and unknown outside.

Another quantity of interest is the *surface* pressure on the vibrator which is determined by one of two methods. The first, used for the FFT approach, is to use Rayleigh's second integral formula

$$p(\mathbf{x}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\mathbf{y}) \frac{\partial}{\partial z_3} \left[\frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|} \right] d\Gamma(\mathbf{y}) \tag{2}$$

and to invert it by the same method used for Eq. (1). The second uses the reconstruction of $\dot{w}(\mathbf{y})$ and Eq. (1) directly moving $\mathbf{x}$ down to the surface ($\mathbf{x}\in\Gamma_p$), which turns out to be more efficient for the SVD approach, as we will discuss below. Given either of these approaches, calculation of the surface pressure along with the reconstructed velocity allows for the computations of the surface normal acoustic intensity ($1/2\Re[p\dot{w}^*]$) and the power radiated from the surface patch.

One fact should be clear from the very beginning: given that d is very small (in the centimeter range)—the inversion of Eq. (1) when the pressure is known only on the patch Ω_p will provide information about $\dot{w}$ only in the area Γ_p directly below this patch, defined by the normal projection of Ω onto the source plane Γ. This fact leads us to the conclusion that one can not determine $\dot{w}(\mathbf{y}\notin\Gamma_p)$ given $p(\mathbf{x}\in\Omega_p)$. In other

words the inversion of Eq. (1) is spatially local with the velocity at a point in Γ_p most effected by the pressure at the normal projection of that point onto Ω_p. This fact leads to the plausibility of patch holography.

There are two distinct methods to invert Eq. (1) and solve for $\dot{w}$. The first is based on the spatial Fourier transform (Fourier acoustics) and the second based on the singular value decomposition (SVD). The SVD approach will be discussed in Sec. II C below.

B. Solution using the FFT

Before we propose the formulation for patch holography we review some theory for FFT based NAH including the recent applications of regularization theory. Transforming Eq. (1), where $\dot{w}$ is given in the infinite domain Γ_∞, using Fourier transforms yields[4]

$$p(\mathbf{x}) = \mathcal{F}^{-1} G_N \mathcal{F} \dot{w}(\mathbf{y}), \tag{3}$$

with $\mathbf{x} \in \Omega_\infty$, $\mathbf{y} \in \Gamma_\infty$. $\mathcal{F}$ represents a two-dimensional Fourier transform operator and $\mathcal{F}^{-1}$ its inverse which obeys $\mathcal{F}^{-1}\mathcal{F} f(x_1,x_2) = f(x_1,x_2)$. We call $\mathcal{F}\dot{w}$ and $\mathcal{F}p$ the k-space spectrum of the velocity and pressure, respectively, where the k-space vector is $\mathbf{k} = (k_1,k_2,k_z)$. The components of Eq. (3) in planar coordinates (x_1,x_2,x_3) are

$$\begin{aligned} &\mathcal{F} \equiv \int \int_{-\infty}^{\infty} dx_1\, dx_2\, e^{-ik_1 x_1} e^{-ik_2 x_2}, \\ &G_N(k_1,k_2) = \frac{\rho c k}{k_z} e^{ik_z d}, \\ &k_z \equiv \sqrt{k^2 - k_1^2 - k_2^2}, \\ &\mathcal{F}^{-1} \equiv \frac{1}{(2\pi)^2} \int \int dk_1\, dk_2\, e^{ik_1 x_1} e^{ik_2 x_2}, \end{aligned} \tag{4}$$

and $x_3 = d$ is the distance between the two parallel surfaces.

In the NAH implementation of Eq. (3) the continuous FT is replaced by the discrete Fourier transform, implemented using the FFT. Furthermore, since the inversion of Eq. (3) is ill-posed we regularize the inversion using a k-space filter $F^\alpha(k_1,k_2)$.[28] The regularization filters out the very high wave number evanescent waves which are due to noise in the measurement. The result is a spatially filtered reconstruction $\tilde{w}$ of the velocity

$$\tilde{w}(\mathbf{y}) = \mathcal{F}^{-1} F^\alpha G_N^{-1} \mathcal{F} p(\mathbf{x}), \tag{5}$$

where

$$G_N^{-1}(k_1,k_2) = \frac{k_z}{\rho c k} e^{-ik_z d} \tag{6}$$

is the inverse propagator in k-space (which increases exponentially as $|k_1| \to \infty$ or $|k_2| \to \infty$), and the algebraic form of the filter is

$$F^\alpha(k_1,k_2) = \frac{|\lambda(k_1,k_2)|^2}{|\lambda(k_1,k_2)|^2 + \alpha \left(\dfrac{\alpha}{\alpha + |\lambda(k_1,k_2)|^2} \right)^2}. \tag{7}$$

In this equation $\lambda(k_1,k_2) \equiv G_N$ is given in Eq. (4) and α is the regularization parameter determined by the Morozov discrepancy principle.[28] Equation (5) requires $p(\mathbf{x} \in \Omega_\infty)$ to solve for $\tilde{\dot{w}}(\mathbf{y} \in \Gamma_\infty)$.

The practical implementation of Eq. (3) entails the use of the DFT to approximate the infinite Fourier transform, discretization of the pressure and velocity apertures, and truncation of the infinite aperture Ω_∞ to a finite aperture extending beyond the vibrator surface Ω_v. This leads[28] eventually to a linear matrix equation if we recast p and $\dot{w}$ as one-dimensional vectors (still spanning two-dimensional space). Thus Eq. (3) is approximated by the linear matrix equation

$$\mathbf{p}(\mathbf{x}) = \mathbf{F}^{-1} \mathbf{G_N} \mathbf{F} \dot{\mathbf{w}}(\mathbf{y}), \tag{8}$$

where column vectors ($\mathbf{p}$ and $\dot{\mathbf{w}}$) and matrix quantities ($\mathbf{G_N}$) are shown in boldface and $\mathbf{G_N}$ is diagonal with elements given by $G_N(k_1,k_2)$. The inverse [Eq. (5)] is

$$\tilde{\dot{\mathbf{w}}}(\mathbf{y}) = \mathbf{F}^{-1} \mathbf{G_N}^{-1} F^\alpha \mathbf{F} \mathbf{p}(\mathbf{x}). \tag{9}$$

$\mathbf{F}$ and $\mathbf{F}^{-1}$ represent the DFT and its inverse represented as matrix multiplies and are implemented using a fast Fourier transform (FFT). Thus $\mathbf{F}^{-1}\mathbf{F} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. It is not necessary to explicitly write a representation for $\mathbf{F}$ as it is easily computed using any standard two-dimensional FFT routine. Also in the implementation of Eq. (9) we must represent $\mathbf{G_N}^{-1}$ [given by Eq. (6)] and F^α [given by Eq. (7)] as diagonal matrices, to implement the simple products in k space called for in Eq. (5). As usual $\mathbf{Fp}$ represents the Fourier coefficients of the discretized pressure in k space.

We will find it very useful later to consider what pressure field is created when the smoothed reconstructed velocity is forward propagated to the hologram plane. That is given the filtered surface velocity $\tilde{\dot{\mathbf{w}}}$ the corresponding filtered pressure field $\tilde{\mathbf{p}}$ is provided by Eq. (8) with no filter term:

$$\tilde{\mathbf{p}} = \mathbf{F}^{-1} \mathbf{G_N} \mathbf{F} \tilde{\dot{\mathbf{w}}}, \tag{10}$$

where the diagonal of $\mathbf{G_N}$ is given as before by G_N of Eq. (4). If we substitute Eq. (9) into this equation we arrive at a simple equation for the filtered (smoothed) pressure in terms of the (unsmoothed) measured pressure $\mathbf{p}$:

$$\tilde{\mathbf{p}} = \mathbf{F}^{-1} F^\alpha \mathbf{F} \mathbf{p}. \tag{11}$$

This equation will be critical for the patch holography algorithm presented in Sec. III.

There is a very important consideration in the reconstruction of the normal velocity given by Eq. (9). We need to compute the regularization parameter α in the filter F^α in Eq. (7). This is accomplished in this case using the Morozov discrepancy principle, although there are other approaches.[28,29] The Morozov discrepancy principle is defined by

$$(\tilde{\mathbf{p}} - \mathbf{p})^H (\tilde{\mathbf{p}} - \mathbf{p}) / M = \sigma^2, \tag{12}$$

with $\tilde{\mathbf{p}}$ given by Eq. (11), where M is the total number of measurement points and σ is the standard deviation of the noise. This very physical principle is based on the fact that a

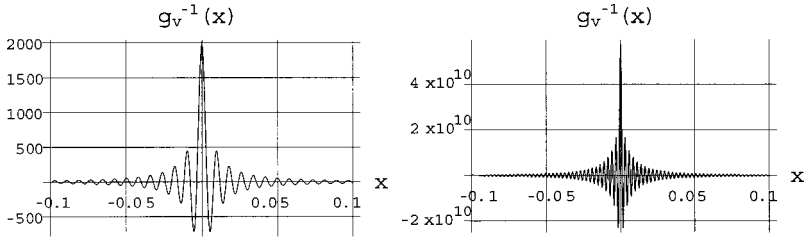


FIG. 2. Inverse real-space velocity propagator (imaginary part) for $k_c=20k$ (left plot) and $k_c=60k$ (right plot), where $k=40$ and $d=0.01$ m. The width of the main lobe in the left plot corresponds to a maximum resolution of about 0.007 m. The peak value at $x=0$ increases towards infinity as k_ρ increases.

perfect filter would completely eliminate the noise from $\mathbf{p}$ yielding $\tilde{\mathbf{p}}-\mathbf{p}=\mathbf{n}$ where $\mathbf{n}$ is a column vector representing the noise at each position in the array. Under these ideal conditions Eq. (12) becomes just the definition of the standard deviation of the noise. Furthermore the latter can be easily determined and is given by the Fourier coefficients corresponding to the average of a set of the smallest spatial wave numbers.[29,28] Note that since $\tilde{\mathbf{p}}$ depends on α then for a given hologram $\mathbf{p}$, Eq. (11) is solved iteratively in conjunction with Eq. (12) until a value of α (and thus $\tilde{\mathbf{p}}$) is found satisfying (to within some tolerance) the latter.[28]

Now consider the pressure reconstruction on the surface of the vibrator. We develop the corresponding linear matrix equation starting with the Fourier transform of Rayleigh's second integral Eq. (2) given by (see Ref. 4)

$$p(\mathbf{x})=\mathcal{F}^{-1}G_D\mathcal{F}p(\mathbf{y}) \quad \text{with} \quad G_D(k_1,k_2)\equiv e^{i\sqrt{k^2-k_1^2-k_2^2}d}, \tag{13}$$

where $p(\mathbf{x})$ is the measured hologram. The transforms in this equation are approximated with the FFT and the pressures represented by column vectors leading to a matrix equation [$\mathbf{G_D}$ a diagonal matrix determined from the equation to the right of Eq. (13)]

$$\mathbf{p}(\mathbf{x})=\mathbf{F}^{-1}\mathbf{G_D}\mathbf{F}\mathbf{p}(\mathbf{y}).$$

and its inverse yielding the reconstructed surface pressure $\overline{\mathbf{p}}(\mathbf{y})$, using the same filter as in Eq. (9),

$$\overline{\mathbf{p}}(\mathbf{y})=\mathbf{F}^{-1}\mathbf{G_D}^{-1}F^\alpha\mathbf{F}\mathbf{p}(\mathbf{x}). \tag{14}$$

With Eq. (9) and Eq. (14) one can compute the active and reactive normal surface intensity.

1. Finite aperture replication problem

As is discussed in Chapter 3 of Fourier acoustics[4] the truncation of the infinite aperture and use of the discrete Fourier transform leads to the aperture replication problem. It is shown there that the degree of approximation inherent in Eq. (9) [and also Eq. (14) by similar arguments] can be modelled with the continuous Fourier transform by the following equation:

$$\tilde{\mathbf{w}}(\mathbf{y})\approx \underbrace{\mathcal{F}^{-1}[F^\alpha G_N^{-1}]}_{g_v^{-1}} ** \sum_{m,n,=-\infty}^{\infty} p_D(x_1-mL_{x_1},x_2-nL_{x_2},d), \tag{15}$$

where

$$p_D(x_1,x_2,d)=\begin{cases} p(x_1,x_2,d) & \text{if } |x_1|\leqslant L_{x_1}/2, |x_2|\leqslant L_{x_2}/2, \\ 0 & \text{otherwise,} \end{cases} \tag{16}$$

** is a two-dimensional spatial convolution. $p(x_1-mL_{x_1},x_2-nL_{x_2},d)$ represents a replicated duplicate of the hologram pressure $p(x_1,x_2,d)$ shifted by the length mL_{x_1} and width nL_{x_2}. Due to Eq. (16) there is no overlapping of the fields and thus the sum in Eq. (15) presents an infinite pressure field composed of replicas of the pressure in the original aperture butted against one another out to infinity in both directions.

The first term in Eq. (15) g_v^{-1} is called the filtered inverse velocity propagator and is singular (much like a Dirac delta function) when there is no filter ($F^\alpha=1$). To understand its behavior we choose a simpler form than Eq. (7) for the filter making it a simple low pass rectangular filter with break point k_c, $F^\alpha=1-H_{-1}(k_\rho/k_c)$ where H_{-1} is the Heaviside step function. This leads to the representative plots along the x_1 axis ($x_2=0$) of the inverse propagator, selecting two values of k_c and shown in Fig. 2. The left plot of Fig. 2 is for $k_c=20k$ and the right for $k_c=60k$ demonstrating that as $k_c\to\infty$ the maximum value becomes infinite and the oscillation frequency increases, very reminiscent of the behavior of the Dirac delta function when defined as a limiting function. The finite spread along the x axis of the inverse propagator leads to the fact that the convolution in Eq. (15) defined by folding $g_v^{-1}(x)$ and shifting it across the replicated apertures picks up an erroneous contribution when it crosses the boundary formed by the adjacent replication of the hologram, thus generating an error localized to this boundary. This error is called the replicated aperture error.

Thus the adjacent (replicated) holograms participate in the reconstruction so that values of $\tilde{\mathbf{w}}(\mathbf{y})$ at the ends of the measurement aperture must be in error due to the contribution of the replica closest to this edge. This edge error is unacceptable for a patch NAH approach. Since the number of measurements points in the hologram is small it is not reasonable to throw out the edge points of $\tilde{\mathbf{w}}(\mathbf{y})$. However, the edge error can be made acceptable by zero-padding the measurement hologram (which automatically zero-pads the replicas) so that the contribution from adjacent replicas can be minimized near the edges. Unfortunately, this zero-padding creates a strong discontinuity which has a very deleterious effect and as a result is not an acceptable solution. That is the filtered inverse propagator in the first term of Eq. (15), in an attempt to keep with the physics of sound propagation, will create an excessively large amplitude, spatially oscillating velocity ($\tilde{\mathbf{w}}(\mathbf{y})$) just below the discontinuity

formed by the zero-padding in an effort to reproduce the discontinuous change of the radiated pressure a small distance away. We will find in Sec. III that the solution to this problem is not to zero-pad but to create a kind of analytic continuation of the hologram measurement which is continuous across the edge boundaries but at the same time matches the actual field outside as closely as possible.

C. Solution using the SVD

The second approach to solving Eq. (1) does not rely on the fast Fourier transform and thus avoids the aperture replication problem. In this method one discretizes the integral equation in real space to produce a matrix equation relating the column vectors p and $\dot{w}$ through a matrix transfer function $\mathbf{G}$:

$$\mathbf{p}=\mathbf{G}\dot{\mathbf{w}}. \tag{17}$$

For notational simplicity assume that $\mathbf{p}$ and $\dot{\mathbf{w}}$ are both of length N so that $\mathbf{G}$ is a $N\times N$ square matrix, although for the procedure presented below the length of $\mathbf{p}$ may be smaller or larger than the length of $\dot{\mathbf{w}}$.

A great deal can be said about the discretization process for Eq. (1), although we do not want to divert attention away from the main purpose of this paper. However, significant effort was expended in developing an accurate discretization scheme so that reconstruction errors due to discretization would be negligible. The main problem arises from the near vanishing of $|\mathbf{x}-\mathbf{y}|$ for the closest points in Ω_p and Γ_p (located on the normal to the surfaces). Although boundary element schemes using shape functions are the usual approach to discretization of the governing integral equations, it proved to be less accurate than a straightforward Taylor series expansion based on the following evaluation of the integral over a square element $\partial\Gamma$ centered at the reconstruction point $\mathbf{y}_0$, with $\mathbf{y}=(y_1,y_2,0)$ and $\mathbf{y}_0=(y_{10},y_{20},0)$:

$$\begin{aligned}
&\int\int_{\partial\Gamma}\dot{w}(\mathbf{y})g(\mathbf{x}|\mathbf{y})d\Gamma(\mathbf{y})\\
&\approx\dot{w}(\mathbf{y}_0)\int\int_{\partial\Gamma}g(\mathbf{x}|\mathbf{y})d\Gamma(\mathbf{y})+\frac{\partial\dot{w}(\mathbf{y}_0)}{\partial y_1}\\
&\quad\times\int\int_{\partial\Gamma}(y_1-y_{10})g(\mathbf{x}|\mathbf{y})d\Gamma(\mathbf{y})+\frac{\partial\dot{w}(\mathbf{y}_0)}{\partial y_2}\\
&\quad\times\int\int_{\partial\Gamma}(y_2-y_{20})g(\mathbf{x}|\mathbf{y})d\Gamma(\mathbf{y})+\frac{\partial^2\dot{w}(\mathbf{y}_0)}{\partial y_1\partial y_2}\\
&\quad\times\int\int_{\partial\Gamma}(y_1-y_{10})(y_2-y_{20})g(\mathbf{x}|\mathbf{y})d\Gamma(\mathbf{y}),
\end{aligned}\tag{18}$$

where $g(\mathbf{x}|\mathbf{y})\equiv e^{ik|\mathbf{x}-\mathbf{y}|}/|\mathbf{x}-\mathbf{y}|$. The length of the sides of the square element were set equal to the lattice spacing of the measurement hologram. The integrations where evaluated using a four-point Gaussian quadrature in each dimension. It should be noted that Kang and Ih have also studied inaccuracy in the BEM approach for NAH applications.[14]

Returning to the reconstruction of $\dot{\mathbf{w}}$ and the linear equation resulting from the Taylor series discretization, the SVD of $\mathbf{G}$ in Eq. (17) is

$$\mathbf{G}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H,\quad \boldsymbol{\Sigma}=\mathrm{diag}(\sigma_1,\ldots,\sigma_{N^2}), \tag{19}$$

where $\mathbf{U}$ and $\mathbf{V}$ are left and right unitary (orthonormal) matrices, respectively, and $\boldsymbol{\Sigma}$ is a diagonal matrix of real singular values. $\mathbf{V}^H$ is the conjugate transpose of $\mathbf{V}$. Since $\mathbf{U}$ and $\mathbf{V}$ are unitary they satisfy the important relations $\mathbf{U}^H\mathbf{U}=\mathbf{I}$ and $\mathbf{V}^H\mathbf{V}=\mathbf{I}$, where $\mathbf{I}$ is the identity matrix and all the matrices are of size $N\times N$. Thus Eq. (17) leads to the equivalent to Eq. (3) and Eq. (8), using the SVD instead of the FFT:

$$\mathbf{p}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H\dot{\mathbf{w}}, \tag{20}$$

with $\mathbf{U}$ playing the role of the inverse Fourier transform and $\mathbf{V}^H$ the role of the forward Fourier transform. $\boldsymbol{\Sigma}$ represents the propagator, similar to G_N, except that it is real and positive and ordered from largest value to smallest value. The inversion of Eq. (20) is straightforward as long as one realizes that due to the evanescent waves the problem is ill-posed. Thus we use the regularization filter given in Eq. (7) and Eq. (9) with $\lambda(k_1,k_2)$ replaced with σ_i (i is the position of the singular value along the diagonal). The inversion is given by

$$\widetilde{\dot{\mathbf{w}}}=\mathbf{V}\boldsymbol{\Sigma}^{-1}F^{\alpha}\mathbf{U}^H\mathbf{p},\quad \boldsymbol{\Sigma}^{-1}=\mathrm{diag}\left(\frac{1}{\sigma_1},\ldots,\frac{1}{\sigma_{N^2}}\right). \tag{21}$$

This equation corresponds to Eq. (9) for the FFT case. As we did in Eq. (11) we can derive an equation for the radiated pressure in Ω resulting from the reconstructed velocity $\widetilde{\dot{\mathbf{w}}}$ when forward-propagated to the hologram surface:

$$\widetilde{\mathbf{p}}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H\widetilde{\dot{\mathbf{w}}},$$

resulting in

$$\widetilde{\mathbf{p}}=\mathbf{U}F^{\alpha}\mathbf{U}^H\mathbf{p}. \tag{22}$$

In identical fashion to the approach for the FFT, the filter parameter α is determined using the Morozov discrepancy principle Eq. (12) and Eq. (22) using a binary search procedure on α until Eq. (12) is satisfied to some specified accuracy.[28]

For reconstruction of the *surface* pressure, unlike the FFT approach, Eq. (2) is not used. Instead we start with the computed reconstruction $\widetilde{\dot{\mathbf{w}}}$, then recompute $\mathbf{G}$ for $\mathbf{x}\in\Gamma$ (on the vibrator surface) and finally compute the surface pressure using Eq. (17):

$$\overline{\mathbf{p}}=\mathbf{G}_{\mathbf{s}}\widetilde{\dot{\mathbf{w}}},$$

where $\mathbf{G}_{\mathbf{s}}$ is the recomputed $\mathbf{G}$ and the bar over $\mathbf{p}$ indicates a surface quantity. In the computation of $\mathbf{G}_{\mathbf{s}}$ we still use Eq. (18). However, to increase the accuracy, we split each of the integrands in Eq. (18) into two terms using

$$\frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|}=\frac{e^{ik|\mathbf{x}-\mathbf{y}|}-1}{|\mathbf{x}-\mathbf{y}|}+\frac{1}{|\mathbf{x}-\mathbf{y}|}$$

and integrate the four Taylor terms containing the $1/|\mathbf{x}-\mathbf{y}|$ term *analytically*. Since the first term is no longer singular we need only one expansion term for the Taylor series and thus we need compute only the first term on the right-hand

side of Eq. (18). This evaluation proved to be extremely accurate. Also because we are solving a forward problem there is no need to compute the SVD of $\mathbf{G_s}$.

The SVD approach has one large advantage over the FFT approach—it does not suffer from the finite aperture replication problem. On the other hand, for a relatively large number of points in the hologram, the SVD approach is orders of magnitude slower. Of course for general geometries the FFT approach can not be used and one must turn to the SVD formulation. Fortunately for the implementation of patch holography the measurement apertures are very small so that the resulting matrices are small and the computation times associated with the SVD and associated matrix operations are not extreme. Thus, as we will find, both the SVD and FFT approaches are viable.

A third approach that uses neither the FFT nor the SVD, but which allows for measurements smaller that the source was presented by Hald for planar geometry.[24,25] In this approach a regularized least squares solution of Eq. (17) is derived by transforming, using Parseval's theorem, the products of spatial Green's functions which make up the regularization operator into easily evaluated *k*-space integrals. This approach also avoids the finite aperture problem discussed above for the FFT application and offers an interesting alternative to the approach presented below for planar geometries.

III. FORMULATION FOR PATCH HOLOGRAPHY

Given the important background of understanding discussed above, we now turn to the main thesis of this paper, a formulation for patch holography. We consider a patch $\Omega_p \in \Omega_v$ much smaller than the vibrator surface as shown in Fig. 1. In other words the pressure field is known only over a small region of the vibrator, and we wish to reconstruct as accurately as possible the normal velocity field in Γ_p, on the normal projection of Ω_p. In this case if one uses the FFT approach the discontinuities due to the finite aperture replication of the pressure hologram discussed above lead to unacceptable errors at the edges. Even with the SVD approach the field reconstructed at the edges of the hologram is not as accurate as the reconstruction in the interior. This inaccuracy results from replacing the infinite limits of integration in the integral equation (1) with finite ones. Thus information from points just outside the aperture are no longer included in Eq. (1) resulting is loss of accuracy around the rim. As discussed above zero-padding minimizes the effect of adjacent replications for the FFT approach but would appear to have no benefit, however, for the SVD approach. However, the resulting discontinuity can have large effects on the velocity reconstruction near the rim for the FFT approach, although the effect is less for the SVD approach, as the results of this paper show. Just how large the effects will be depends in a complicated manor on the standoff distance and the maximum spatial wave number resolved in the reconstruction (set by the filter cutoff) although there do not appear to be any simple guidelines that one can use. For the FFT case Hayek[30] provides some good examples including several patch examples for planar geometry.

Thus to avoid the discontinuities we formulate a method to compose or analytically continue the pressure field into a rim Ω_c (see Fig. 1) outside of the actual measurement, based on the field inside. This continuation uses the spectral information about the field in Ω_p and smoothly continues and tapers into Ω_c. The spectral information helps provide a continued field which is not unlike the actual field (unknown however) outside of Ω_p. This creates a synthetic measurement aperture larger than the original, removing the original discontinuity at the edges of Ω_p. Any adjacent replicated aperture discontinuities (if a FFT approach is used) are moved to the outside boundary of Ω_c sufficiently beyond the boundary of the original patch Ω_p so that they do not effect the data reconstructed on the surface in Γ_p.

The method for aperture continuation presented here is similar to one published recently by Saijyou and Yoshikawa[26] and extended by Williams.[27] Saijyou does not provide a SVD approach, dealing only with the FFT approach and also does not apply regularization theory using a simple rectangular low-pass filter instead of Eq. (7) with cutoff value determined *a priori*. The extension by Williams added a regularization filter in the algorithm with automated selection of the regularization parameter α (determines the cutoff).

We describe briefly this extension and give some examples of using the resulting continued field to reconstruct the normal velocity on a patch within the boundary of a vibrating plate, not discussed in Ref. 27. Either the FFT or SVD methods may be used in this approach, the former based on Eq. (11) and the latter on Eq. (22). Define measurement Ω and reconstruction Γ surfaces: $\mathbf{y} \in \Gamma_p \cup \Gamma_c$ and $\mathbf{x} \in \Omega_p \cup \Omega_c$ where the domains Ω_c and Γ_c represent a banded area of fixed width circumscribing Ω_p and Γ_p, respectively. As yet the field is unknown in the "picture frame" regions designated by the subscript *c* and the objective is to continue the pressure field into Ω_c. Figure 1 shows the definitions of the domains.

The extrapolation or continuation of the measured pressure field into the region Ω_c is accomplished using an iteration procedure. Rewrite Eq. (11) and Eq. (22) for the *i*th iteration as

$$\tilde{\mathbf{p}}^{(i)} = \mathbf{H}^{(i)}\mathbf{p}^{(i)} \begin{cases} \mathbf{H}^{(i)} \equiv \mathbf{F}^{-1} F^{\alpha^{(i)}} \mathbf{F} & \text{(FFT)}, \\ \mathbf{H}^{(i)} \equiv \mathbf{U} F^{\alpha^{(i)}} \mathbf{U}^{H} & \text{(SVD)}, \end{cases} \tag{23}$$

where $\mathbf{H}^{(i)}$ indicates that $\mathbf{H}$ changes from iteration to iteration due to the changing filter factor $F^{\alpha^{(i)}}$ whereas the nonsuperscripted terms do not.

The filter factor is given by Eq. (7) with α replaced with $\alpha^{(i)}$, using either $\lambda(k_1,k_2)$ for the FFT or σ_k (diagonal of $\mathbf{\Sigma}$) given in Eq. (19). Note that the singular values do not change throughout the iteration process, only $\alpha^{(i)}$ changes. As described after Eq. (12) the Morozov discrepancy principle is used to determine the value of $\alpha^{(i)}$ using

$$\sigma^{(i)} = \sqrt{(\tilde{\mathbf{p}}^{(i)} - \mathbf{p}^{(i)})^H (\tilde{\mathbf{p}}^{(i)} - \mathbf{p}^{(i)})/M}, \tag{24}$$

and $\sigma^{(i)}$ is the standard deviation of the noise determined from $\|\mathbf{U}^H \mathbf{p}^{(i)}\|$ where the L2 norm is taken using only the last

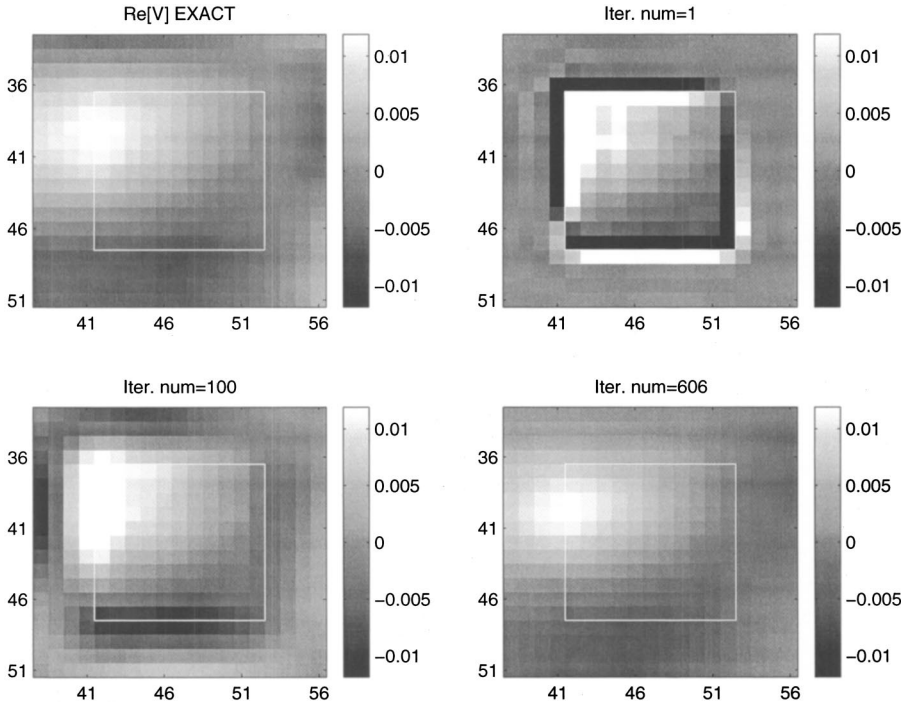


FIG. 3. Surface normal velocity reconstructions for different iterations for the FFT approach. The frequency was 267 Hz. The white outline indicates the boundary of the patch hologram (Γ_p corresponding to Ω_p) located in the interior. Exterior to the outline are Ω_c and Γ_c which represent a band 20 points wide although only a four points band is shown in the figure.

Q rows of $\mathbf{U}^H$ as described in the literature.[28] In this study we used $Q=N^2/4$.

Let the column vector $\mathbf{p}$ (with no superscript) represent the measured pressure data in the patch Ω_p. The iteration procedure is started with the measured pressure field continued into Ω_c by zero-padding:

$$\mathbf{p}^{(0)} \equiv \begin{cases} \mathbf{p}(\mathbf{x}\in\Omega_p) & \text{(measurement)}, \\ 0(\mathbf{x}\in\Omega_c) & \text{(zero-padding)}. \end{cases} \tag{25}$$

Using Eq. (23) the first filtered estimate $\tilde{\mathbf{p}}^{(0)}$ of the continued pressure field is obtained. The filtered estimate is not null in Ω_c and differs from the measured field in Ω_p. We remedy the latter by replacing the estimate in Ω_p with the original measured field without changing the estimate in Ω_c. This procedure defines the next iterate $\mathbf{p}^{(1)}$ as well as future iterates (with $i=0$ initially):

$$\mathbf{p}^{(i+1)} \equiv \begin{cases} \mathbf{p}(\mathbf{x}\in\Omega_p) & \text{(measurement)}, \\ \tilde{\mathbf{p}}^{(i)}(\mathbf{x}\in\Omega_c). \end{cases} \tag{26}$$

The original measured pressure is always used to construct the field in Ω_p for all iterates. With this construction Eq. (23) is used again to compute the continuation $\tilde{\mathbf{p}}^{(i+1)}$ for the next iteration. This process is repeated until some desired convergence $\epsilon(\|\tilde{\mathbf{p}}^{(i+1)}-\tilde{\mathbf{p}}^{(i)}\|<\epsilon)$ is reached. We choose a value $\epsilon = 0.1\sigma^{(i)}\sqrt{M}$, although the actual value is not critical.

Note that the iteration equations do not make explicit use of the reconstructed surface velocity, although information is implicitly encoded in the singular values used for the regularization filter. The surface velocity is determined separately using

$$\tilde{\dot{\mathbf{w}}}^{(i)} = \begin{cases} \mathbf{F}^{-1}\mathbf{G}^{-1}F^{\alpha^{(i)}}\mathbf{F}\tilde{\mathbf{p}}^{(i)} & \text{FFT Eq. (9)}, \\ \mathbf{V}\boldsymbol{\Sigma}^{-1}F^{\alpha^{(i)}}\mathbf{U}^H\tilde{\mathbf{p}}^{(i)} & \text{SVD Eq. (21)}. \end{cases} \tag{27}$$

One can also reconstruct the *surface* pressure $\overline{\mathbf{p}}(\mathbf{x}\in\Gamma_p\cup\Gamma_c)$ using

$$\overline{\mathbf{p}}^{(i)} = \begin{cases} \mathbf{F}^{-1}\mathbf{G}_\mathbf{D}^{-1}F^{\alpha^{(i)}}\mathbf{F}\tilde{\mathbf{p}}^{(i)} & \text{FFT Eq. (14)}, \\ \mathbf{G}_\mathbf{s}\tilde{\dot{\mathbf{w}}}^{(i)} & \text{SVD Eq. (17)}, \end{cases} \tag{28}$$

where $\mathbf{G}_\mathbf{s}$ computed for the pressure field in the same plane as the velocity field by the method discussed in Sec. II C. In all of the five equations above the SVD is calculated only once.

There is an important consideration with respect to the width of the "frame" region. As discussed above for the FFT approach convolution of the replications of the hologram aperture demonstrated in Eq. (15) implies that adjacent holograms [for example, the four holograms given by $(m,n)=(1,0)$, $(0,1)$, $(-1,0)$, $(0,-1)$] cause reconstruction errors in Γ_c, especially at the outer edges of Γ_c. Since $p(x_1,x_2,d)$ in Eq. (16) includes the "frame" Γ_c we can reduce the effects of the adjacent holograms by making the width of Γ_c large. For the case presented here we found that a width of 20 lattice points was sufficient to move the replicated apertures far enough away for an accurate continuation of the field into Ω_c. However, in the case of the SVD there are no replicated holograms, and experience has shown that one can choose the width to be much smaller with two to five points being sufficient. These widths have been used in experiments with standoff distances ranging from 0.4 cm to 3.0 cm with success. Larger standoffs may lead to different results, however.

We now are in a position to study the accuracy of the surface velocity and pressure reconstructions resulting from the continuation of the measurement hologram into Ω_c using the five equations above. We use experimental data for $\mathbf{p}$.

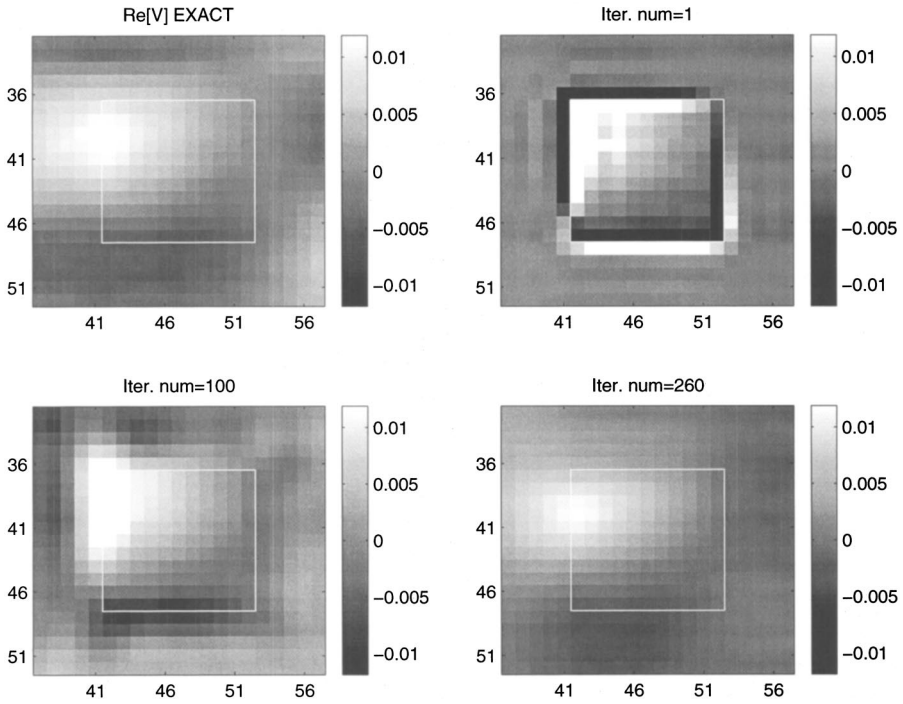


FIG. 4. Surface normal velocity reconstructions from the SVD for different iterations. The frequency was 267 Hz. The white outline indicates the boundary of the patch hologram (Γ_p corresponding to Ω_p). In this case Ω_c and Γ_c correspond to a band five points wide.

A. Surface velocity reconstruction example for FFT based patch NAH

The resulting velocity reconstructions using this iteration procedure are shown in Fig. 3 for the experimental data described in Sec. III C below. Top left plot shows the exact result derived from a complete hologram (69×46 points) extending past the boundaries of the vibrator. The region directly over the plate vibrator (vibrator represented by Γ_v in Fig. 1) corresponded to an array of 45×21 points. A small 11×11 patch Ω_p centered above the plate is then extracted from this complete hologram and used for $\mathbf{p}$ in the iteration algorithm described above. The second plot shows the velocity $\tilde{\mathbf{w}}^{(0)}$ from Eq. (27) using the 11×11 patch after one iteration, corresponding to a reconstruction of the zero-padded (20 point frame width) patch pressure. This plot shows the deleterious large amplitude, spatially oscillating field at the edges due to the zero-padding. The lower left plot, the result after 100 iterations, shows that the spurious velocity field has been reduced. The lower right plot shows the final result after convergence—the spurious velocity field is gone. The velocity errors in Γ_p (area within outlined white box) after 1, 100, and 606 iterations were 149%, 37%, and 8%, respectively.

Another important result shown in Fig. 3 is the fairly accurate reconstruction of the normal velocity field in the region Γ_c outside the measured data, especially close to the boundary (white box) of Γ_p. This accuracy, however, degrades with distance from the boundary. (Although not shown, the reconstructed velocity field generally tapers to zero as the outer edge of Γ_c is reached.)

Now consider the pressure reconstruction $\overline{\mathbf{p}}(\mathbf{y})$ on the surface of the vibrator. This is also a byproduct of the iteration procedure and is given by the upper equation of Eq. (28) for the FFT procedure. It turns out that after convergence the resulting errors for the surface pressure are much smaller that the velocity errors. For the case represented in Fig. 3 the reconstructed pressure $\overline{\mathbf{p}}(\mathbf{y}\in\Gamma_p)$ errors were 13.5%, 3.4%, and 0.7% corresponding to 1, 100, and 606 iterations, respectively. Thus the final error was 0.7% compared with 8% for the normal velocity reconstruction. We found that it was always the case that the reconstructed pressure field was considerably more accurate than the normal velocity field on the vibrator surface.

B. Surface velocity reconstruction example for SVD based patch NAH

The iteration procedure for the SVD is identical in all respects except one to the FFT procedure. For the SVD the region Γ_c and corresponding Ω_c is much smaller. This results from one of the major differences between the SVD and FFT approaches. As discussed above the FFT creates an infinite set of replications of the hologram aperture butted together so that the width of Ω_c should be large enough to move these replications far enough apart to minimize their influence close to Ω_p. We used a width of 20 lattice points. However, for the SVD there is no analog to the replicated holograms and one is free to choose the width of the frame much smaller. In this case we used a five point width although two points is often sufficient. The results for the same conditions as Fig. 3 are shown in Fig. 4. As can be seen the converged results of the FFT (Fig. 3) versus the SVD approaches (Fig. 4) shown in the bottom right plot of each figure are almost identical. Again the discontinuity due to zero-padding appears as a strong velocity field shown in the top right—not surprising since one would expect a pathological velocity field is needed to generate a sudden vanishing (spatially) of the pressure in the plane above. The velocity errors associated with the SVD method were 142%, 31%, and 15% corresponding to one, 100 and 260 iterations,

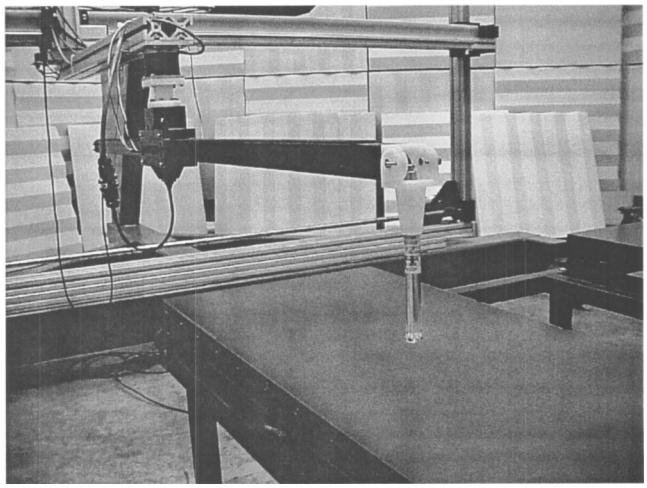


FIG. 5. Steel plate with microphone near its center and the automated $x-y$ scanner shown in the background.

slightly worse than the FFT case. We will compare the FFT and SVD approaches more extensively in the next section.

For the case of the pressure reconstructions on the surface of the vibrator we use the lower equation of Eq. (28). Like the FFT case we again find that the errors in the reconstruction of the surface pressure are much smaller than for the reconstructed surface velocity, in this case being 13%, 2.9%, and 1.3% corresponding to one, 100 and 260 iterations for the result shown in Fig. 4. Thus for the SVD approach the final (after 260 iterations) pressure error was 1.3% compared to the final velocity error of 15%.

C. Experimental results

The experimental configuration for the holographic measurement is shown in Fig. 5. The steel rectangular plate has overall dimensions of 55.9×26.7 cm and a thickness of 1.6 mm. The plate was driven by a Wilcoxon F3/F9 inertial shaker (0 to 3000 Hz) fixed in the underneath corner of the plate in order to generate a broad spectrum of spatial wave numbers. The resulting applied force was used to normalize the pressure measured with a ACO 7046 1/2 inch microphone shown above the plate. The scan is conducted in an automated point-by-point fashion. The microphone is initially positioned by the $x-y$ scanner at a corner of the scan area, where data is then acquired. Next, the microphone is moved to an adjacent point, and data is again acquired. This process is continued until the full scan area is measured. Synchronization between the source shaker, the data acquisition and the scanner has proven to have high stability, yielding coherent data that can be thought of as having been acquired simultaneously. The pressure was measured on a grid of 69 points along the longer dimension and 46 points along the shorter dimension in a plane 0.4 cm above the plate surface. The step size in both the x and y directions was 1.25 cm, providing an overall scan dimension of 85.0×57.5 cm. The (smaller) plate surface is located at the center of this area.

In the results which follow the ''exact'' velocity and surface pressure on the plate were determined by a standard NAH reconstruction using the full 69×46 point array of points which includes a significant over-scan (conformal to Γ_∞ in Fig. 1) of the vibrating area (Γ_v). (The array of points directly over the plate area conformal to Γ_v corresponded to a 45×21 point array.) It has been demonstrated that the reconstructions with the over-scanned hologram are extremely accurate.[4]

The patch iteration algorithm was tested using a small subset of this array—a 10×10 point patch with center coincident with the center of the vibrating plate. Thus the patch size was a 11.3 cm square area. For this study we choose 250 frequencies spaced equally throughout the range of 247 to 2678 Hz and computed the error in Γ_p. As above the FFT used a frame region (Ω_c) 20 points wide, and the SVD a five point wide boundary. Wider frame regions did not improve the results significantly.

The results for the normal velocity reconstruction using the patch iteration algorithm discussed above are shown for both the FFT and SVD approaches in Fig. 6. In the results the normal velocity was normalized by the applied force at each frequency to give the mobility. The exact result (NAH mobility reconstruction using the full 69×46 array of points with mobility root-mean-square averaged over Γ_p) is the solid line, the circle is the FFT and the plus the SVD approach. For each symbol on the plot a patch reconstruction was carried out using the iteration procedure as demonstrated in Figs. 3 and 4. The root-mean-square average mobility over Γ_p (at a single frequency) was computed in dB for each and plotted in the figure. For example, for the SVD approach the column vector $\tilde{\dot{\mathbf{w}}}(\mathbf{x}\in\Gamma_p)$ represents the reconstructed velocity per unit force (mobility) and the RMS spatial average in dB plotted in the figure is given by $20\log_{10}(\tilde{\dot{\mathbf{w}}}^H\tilde{\dot{\mathbf{w}}})^{1/2}$. The spatially averaged mobility display is particularly useful as it also demonstrates the resonances (peak mobilities) and anti-resonances (valleys) of the point-driven plate. As can be seen from the figure there is generally less than one dB difference between the SVD, FFT and exact results over the full frequency range. The results are excellent.

A quantitative measure of the error is also provided in the figure, in the legend in the upper right-hand corner. This error is a simple 250 point average of the error-per-frequency $E(\omega)$ determined using

$$E(\omega)=100\left(\frac{(\tilde{\dot{\mathbf{w}}}-\dot{\mathbf{w}})^H(\tilde{\dot{\mathbf{w}}}-\dot{\mathbf{w}})/N_{\Gamma_p}}{\dot{\mathbf{w}}_\mathbf{t}^H\dot{\mathbf{w}}_\mathbf{t}/N_t}\right)^{1/2}, \tag{29}$$

where the numerator is divided by the number of points in Γ_p, in this case $N_{\Gamma_p}=100$ (length of $\dot{\mathbf{w}}$), and the denominator is divided by the total number of points $N_t=945$ (length of $\dot{\mathbf{w}}_\mathbf{t}$) on the total plate surface Γ_v. In this formula the exact velocity reconstruction (per unit force) over Γ_p is the vector $\dot{\mathbf{w}}$ and the exact velocity reconstruction (per unit force) over the whole plate is the vector $\dot{\mathbf{w}}_\mathbf{t}(\dot{\mathbf{w}}\in\dot{\mathbf{w}}_\mathbf{t})$ both determined from the full hologram (3105 points) which over-scanned the whole plate. Thus in this error formula, the denominator can be viewed as the RMS (root-mean-square) average velocity field per unit force over the complete surface of the vibrator, and the numerator the RMS average velocity field *difference* per unit force only over the patch. Finally, the simple average

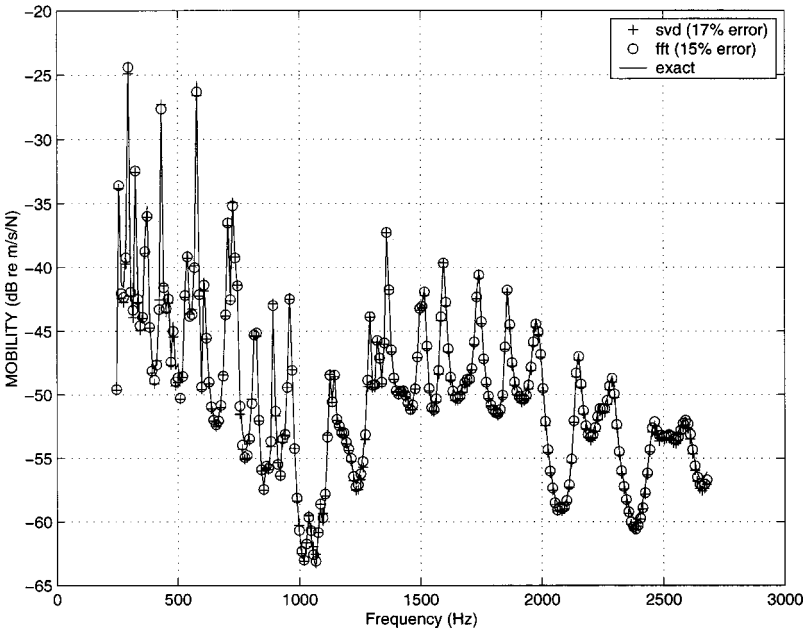


FIG. 6. Results for the RMS spatial average over the patch of the normal velocity reconstruction normalized by the driving force (mobility) using the patch algorithm. The SVD approach is shown with + and the FFT with ○. The exact result is the solid line.

of $E(\omega)$ (the ''summary error'') over all the frequencies is indicated in the legend in the plot. For the FFT approach the ''summary error'' was 15%, and for the SVD 17%. It is interesting to note that the errors using just the center 36 values of $\tilde{\dot{\mathbf{w}}}-\dot{\mathbf{w}}(N_{\Gamma_p}=36)$ in the numerator of Eq. (29) was 6.3% for the FFT and 6.9% for the SVD. This indicates that the errors increase as one approaches the edge of the hologram patch.

As discussed above the reconstruction of the surface pressure on Γ_p, not shown here, was very accurate with appreciably smaller errors levels than the normal velocity reconstructions. A more difficult quantity to reconstruct is the normal acoustic intensity in Γ_p. This requires a reconstruction of both the surface pressure and velocity and computation of the active intensity from the real part of one-half the product of the pressure and the complex conjugate of the velocity. Intensity results are shown in Fig. 7 with the same symbol definitions as the previous figure. The average errors across the band are low, 27% for the FFT and 30% for the SVD. The corresponding errors for the center 36 points were 13% and 15%, respectively. The agreement with ''truth'' is very good and one would conclude that patch NAH can be used to determine the surface normal intensity (and thus the total power radiated from the patch region) with reasonable success. However, the ''truth,'' based on a classic NAH reconstruction from the complete 69×46 point array, is subject to error itself especially at the resonances of the plate (where the pressure and velocity move closer to being 90 degrees apart).

Finally, to demonstrate the improvement which is gained from using the ''analytic continuation'' method which forms the backbone of the patch NAH algorithm, we provide results of a NAH patch reconstruction which uses no extension or continuation of the measurement aperture data (regions

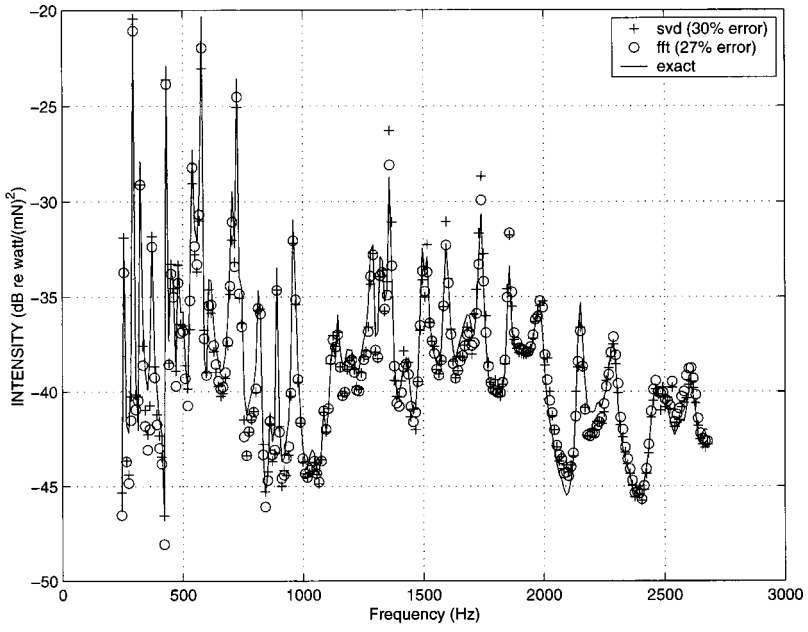


FIG. 7. Results for the normal (active) intensity reconstruction normalized by the driving force using the patch algorithm. The SVD approach is shown with + and the FFT with ○. The ''exact'' result is the solid line, although in this case it too is subject to some error due to noise in the data, even though the full hologram is used.

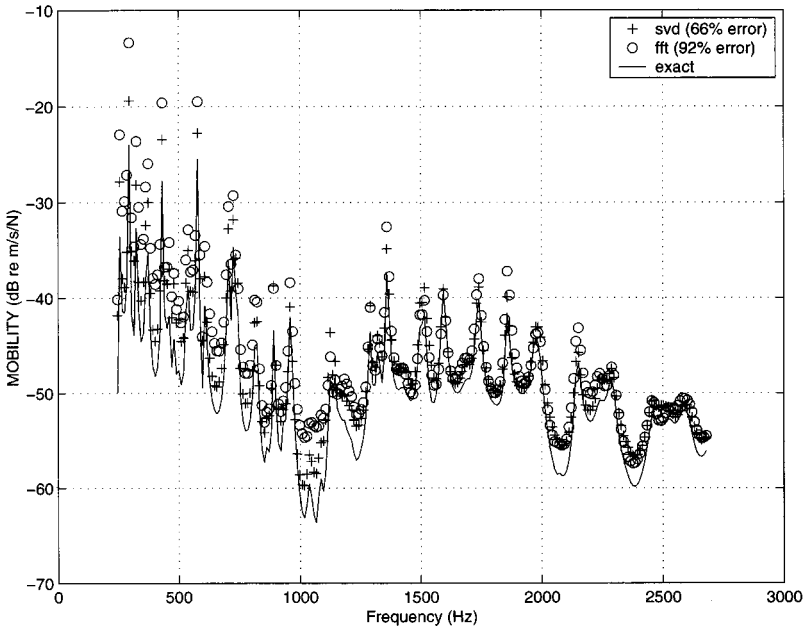


FIG. 8. The result if no ''analytic continuation'' of the hologram is used (region Ω_c is null). Results for the normal velocity reconstruction normalized by the driving force using NAH on the same 10×10 patch as before but with no zero-padding and no iteration.

Ω_c and Γ_c have zero width). The reconstruction results for the averaged surface mobilities are shown in Fig. 8. The errors have increased, compared with Fig. 6, from 17% to 66% for the SVD and from 15% to 92% for the FFT. It is not surprising that the SVD now does better because it does not suffer from the finite aperture replication discontinuity problem inherent in the FFT approach. It is notable, however, that the SVD errors also increased substantially, even though there is no replication problem with the SVD formulation. This error increase is due to the truncation of the propagator matrix **G** which prevents adjacent points just outside of Γ_p from contributing to the inversion of **G** thus increasing the errors for points near the rim versus the interior. If we again look at the errors in just the center region of 36 points the errors for the SVD reduced from 66% to 24% and the FFT from 92% to 35% illustrating the fact that the edge problems in both these methods are very local.

IV. CONCLUSIONS

Although we have demonstrated the success of patch NAH only for planar geometry, the use of the SVD allows immediate extension to any other geometry, given that one has a boundary element code available to generate the Green matrix **G**. The only requirement is that the two surfaces Γ_p and Ω_p should remain parallel or conformal to one another. This guarantees that the evanescent waves decay uniformly at all points from Γ_p to Ω_p.

We have demonstrated that the ''analytic continuation'' of the measured pressure field leads to significantly improved reconstructions of the normal velocity over a simple truncated aperture approach. This ''analytic continuation'' forms the cornerstone of a patch NAH approach which can be applied to simple geometries using the FFT or more complex geometries by using the SVD. The pressure, velocity, and intensity on a patch of a vibrating surface is accurately determined by this approach. This greatly simplifies the measurement requirements of NAH in experimental applications in which only part of a vibrator is of interest. This added flexibility provides a new front of attack for NAH in problems which were not experimentally feasible before.

Furthermore the need for fewer measurement points opens up the possibility of using conventional two-dimensional microphone arrays for instantaneous measurements, expanding the capability of NAH to reconstruct multiple random sources which may be nonstationary and incoherent.

ACKNOWLEDGMENTS

Work supported by the Office of Naval Research.

[1] E. G. Williams and J. D. Maynard, ''Holographic imaging without the wavelength resolution limit,'' Phys. Rev. Lett. **45**, 554–557 (1980).
[2] S. Yoshikawa, E. G. Williams, and K. B. Washburn, ''Vibration of two concentric submerged cylindrical shells coupled by the entrained fluid,'' J. Acoust. Soc. Am. **95**, 3273–3286 (1994).
[3] K. Saijyou, M. Ueda, and S. Yoshikawa, ''Applicaton of generalized nearfield acoustical holography to scattering problems,'' Jpn. J. Appl. Phys., Part 1 **33**, 3170–3175 (1994).
[4] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London, UK, 1999).
[5] J. Hald, ''Stsf–a unique technique for scan-based nearfield acoustical holography without restriction on coherence,'' Technical Report 1, B&K Rechnical Review, 1988.
[6] W. A. Veronesi and J. D. Maynard, ''Digital holographic reconstruction of sources with arbitrarily shaped surfaces,'' J. Acoust. Soc. Am. **85**, 588–598 (1989).
[7] G. Borgiotti, A. Sarkissian, E. G. Williams, and L. Schuetz, ''Generalized nearfield acoustic holography for axisymmetric geometries,'' J. Acoust. Soc. Am. **88**, 199–209 (1990).
[8] G.-T. Kim and B.-H. Lee, ''3-D sound source reconstruction and field repreediction using the helmholtz integral equation,'' J. Sound Vib. **136**, 245–261 (1990).
[9] A. Sarkissian, ''Nearfield acoustical holography for an axisymmetric geometry: a new formulation,'' J. Acoust. Soc. Am. **88**, 961–966 (1990).
[10] M. R. Bai, ''Application of bem (boundary element method)-based acoustic holography to radiation analysis of sound sources with arbitrarily shaped geometries,'' J. Acoust. Soc. Am. **92**, 533–549 (1992).
[11] B.-K. Kim and J.-G. Ih, ''On the reconstruction of the vibro-acoustic field over the surface enclosing an interior space using the boundary element method,'' J. Acoust. Soc. Am. **100**, 3003–3016 (1996).
[12] E. G. Williams, B. H. Houston, P. C. Herdic, S. T. Raveendra, and B.

Gardner, ''Interior nah in flight,'' J. Acoust. Soc. Am. **108**, 1451–1463 (2000).
[13] B.-K. Kim and J.-G. Ih, ''Design of an optimal wave-vector filter for enhancing the resolution of reconstructed source field by nah,'' J. Acoust. Soc. Am. **107**, 3289–3297 (2000).
[14] S.-C. Kang and J.-G. Ih, ''Use of nonsingular boundary integral formulation for reducing errors due to near-field measurements in the boundary element method based near-field acoustic holography,'' J. Acoust. Soc. Am. **109**, 1320–1328 (2001).
[15] Z. Zhang, N. Vlahopoulos, T. Allen, and K. Y. Zhang, ''A source reconstruction process based on a indirect variational boundary element formulation,'' Eng. Anal. Boundary Elem. **25**, 93–114 (2001).
[16] Z. Wang and S. F. Wu, ''Helmholtz equation-least-squares method for reconstructing the acoustic pressure field,'' J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[17] S. F. Wu and J. Yu, ''Reconstructing interior acoustic pressure field via helmholtz equation least-squares method,'' J. Acoust. Soc. Am. **104**, 2054–2060 (1998).
[18] E. G. Williams, ''On green functions for a cylindrical cavity,'' J. Acoust. Soc. Am. **102**, 3300–3307 (1997).
[19] F. Augusztinovicz, ''Reconstruction of source strength distribution by inversing the boundary element method,'' in *Boundary Elements in Acoustics, Advances and Applications*, edited by Otto von Estorff (WIT Press, Southhampton, UK, 2000).
[20] P. A. Nelson and S. H. Yoon, ''Estimation of acoustic source strength by inverse methods: Part i, conditioning of the inverse problem,'' J. Sound Vib. **233**, 643–668 (2000).
[21] S. H. Yoon and P. A. Nelson, ''Estimation of acoustic source strength by inverse methods: Part ii, experimental investigation of methods for choosing regularization parameters,'' J. Sound Vib. **233**, 669–705 (2000).
[22] S. H. Yoon and P. A. Nelson, ''A method for the efficient consturcution of acoustic pressure cross-spectral matrices,'' J. Sound Vib. **233**, 897–920 (2000).
[23] A. Sarkissian, C. Gaumond, E. Williams, and B. Houston, ''Reconstruction of the acoustic field over a limited surface area on a vibrating cylinder,'' J. Acoust. Soc. Am. **93**, 48–54 (1993).
[24] R. Steiner and J. Hald, ''Near-field acoustical holography without the errors and limitations caused by the use of spatial DFT,'' Int. J. Acoust. Vib. **6**, 83–89 (2001).
[25] J. Hald, ''Planar near-field acoustical holography with arrays smaller than the sound source,'' Proceedings of the 17th International Congress on Acoustics, Rome, Italy, 2001, Vol. I, part A.
[26] K. Saijyou and S. Yoshikawa, ''Reduction methods of the reconstruction error for large-scale implementation of near-field acoustical holography,'' J. Acoust. Soc. Am. **110**, 2007–2023 (2001).
[27] E. G. Williams, ''Continuation of acoustic near-fields,'' J. Acoust. Soc. Am. **113**, 1273–1281 (2003).
[28] E. G. Williams, ''Regularization methods for near-field acoustical holography,'' J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[29] Per Christian Hansen, *Rank-Deficient and Discrete Ill-Posed Problems* (Siam, Philadelphia, PA, 1998).
[30] S. I. Hayek and T. W. Luce, ''Aperture effects in planar nearfield acoustical imaging,'' Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **110**, 91–96 (1988).

A dynamic Young's modulus measurement system for highly compliant polymers

François M. Guillot[a] and D. H. Trivett
George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 6 May 2003; accepted for publication 7 July 2003)

A new system to determine experimentally the complex Young's modulus of highly compliant elastomers at elevated hydrostatic pressures and as a function of temperature is presented. A sample cut in the shape of a bar is adhered to a piezoelectric ceramic shaker and mounted vertically inside a pressure vessel equipped with glass windows. Two independent measurement methods are then used: a resonant technique, to obtain low-frequency data, and a wave propagation technique, to obtain higher-frequency data. Both techniques are implemented utilizing laser Doppler vibrometers. One vibrometer detects sample resonances through a window located at the bottom of the pressure vessel, and a set of two separate vibrometers monitors the speed of longitudinal waves propagating in the sample, through windows located on the sides of the vessel. The apparatus is contained inside an environmental chamber for temperature control. Using this approach, Young's modulus data can be obtained at frequencies typically ranging from 100 Hz to 5 kHz, under hydrostatic pressures ranging from 0 to 2.07 MPa (300 psi), and at temperatures between −2 °C and 50 °C. Experimental results obtained on two commercial materials, Rubatex® R451N and Goodrich Thorodin™ AQ21, are presented. The effects of lateral inertia, resulting in dispersive wave propagation, are discussed and their impacts on Young's modulus measurements are examined. [DOI: 10.1121/1.1604121]

PACS numbers: 43.20.Ye, 43.20.Jr, 43.20.Ks [YHB]

I. INTRODUCTION

Compliant polymers, such as voided polyurethanes, are materials of interest because of their vibration isolation and damping properties. Knowledge of the dynamic elastic properties of such elastomers, as a function of temperature and hydrostatic pressure, is essential to predict their performance, to validate numerical models, and to design sonar systems. As is the case for any homogeneous, isotropic material, only two elastic constants (along with the density) are needed to completely characterize them. Among these elastic constants, the bulk and Young's moduli are the easiest to obtain experimentally, and also constitute the best pair, in terms of computational error minimization, from which to compute other moduli. In the Acoustic Material Laboratory of the Georgia Institute of Technology, two independent experimental systems have been developed to perform dynamic measurements of these two moduli. The present article describes the dynamic Young's modulus apparatus and presents data that illustrate a portion of the capabilities of the measurement system. These data were obtained on a commercial closed-cell foam neoprene, Rubatex® R451N, and on a commercial elastomer, Goodrich Thorodin™ AQ21.

Several methods have been used to dynamically measure the Young's modulus of viscoelastic materials. The most widely employed approach is the resonant bar technique introduced by Norris and Young,[1,2] where a sample in the shape of a bar is excited harmonically at one end and the ratio of the end accelerations is measured. At resonance, the response of the sample can be used to obtain values of the complex modulus. Madigoski and Lee[3,4] utilized the technique extensively to characterize a number of materials as a function of temperature, and used the time–temperature superposition principle[5] (or WLF shift procedure, named after Williams, Landel, and Ferry) to obtain data over extended frequency ranges. Recently, the resonance method has been adopted as a standard by the American National Standard Institute.[6] Garrett[7] used an electrodynamic transduction scheme to excite torsional, longitudinal, and flexural modes in rods of circular or elliptical cross section. Both the longitudinal and flexural modes yield values for the Young's modulus. Garrett measured the resonance frequencies of these modes to compute values of the magnitude of the modulus. Guo and Brown[8] extended this method to measure the complex Young's modulus by fitting the analytical solution of the longitudinal wave equation to experimental data. Their approach allows the determination of the modulus at frequencies that are not restricted to the resonance frequencies of the modes. Measurements as a function of both temperature and hydrostatic pressure were reported by Willis *et al.*,[9] between 7 °C and 40 °C and over a 0- to 3.45-MPa (500-psi) range. Their approach consisted of measuring, by laser vibrometry, the dynamic response of a rectangular block excited by a shaker, and matching the data with predictions from a finite-element code in which the complex elastic moduli were the adjustable parameters. Their sample was inside a pressure chamber submerged in a water bath for temperature control. Although this method allows one to obtain data over relatively large frequency ranges without the

[a]Author to whom correspondence should be addressed. Electronic mail: francois.guillot@me.gatech.edu

need for time–temperature shifts, it suffers from occasional convergence problems associated with the inversion of the finite-element code.

The method described in this article uses an improved version of the resonant bar method and combines it with a wave-speed measurement technique. These approaches, which are implemented using laser Doppler vibrometers, have the advantage of providing *directly measured* complex wave-speed data (which are converted to Young's modulus data, using the density of the material) over a much broader frequency range than that typically afforded by the resonant bar method alone. Therefore, fewer temperature data sets are needed in order to generate master curves, reducing the amount of uncertainty associated with multiple WLF shifts. Furthermore, since all the sample displacements are measured with noncontact vibrometers, no mass corrections (due to the presence of accelerometers or electromagnetic transducers) need to be taken into account. Samples are placed inside a stainless-steel vessel which is itself contained inside an environmental chamber, providing pressure and temperature capabilities over a 0- to 2.07-MPa (300-psi) range and a $-2\,^{\circ}\mathrm{C}$ to $50\,^{\circ}\mathrm{C}$ range, respectively.

II. THEORETICAL FRAMEWORK

The theoretical background of the techniques employed in the system can be compiled from several articles,[1,3,4,8] technical reports,[2,10] and textbooks.[11–14] However, because some of these references are not readily available and because none of them provides, in one single convenient place, all the information relevant to this work, we chose to present below the most important theoretical points of the measurements.

A. Resonance measurements

Consider the forced vibrations of a homogeneous bar of density ρ and length L, with a constant cross section and no mass attached to its free end (Fig. 1). One end of the bar is driven with a harmonic displacement $u_0(t)=U_0e^{i\omega t}$, and the resulting axial displacement (relative to the driven end) at a distance x from the driven end is $u(x,t)=U(x)e^{i\omega t}$. Assuming a uniform, uniaxial stress distribution inside the bar, and neglecting the effects of lateral inertia, the equation of motion can be written as

$$\frac{\partial\sigma_x}{\partial x}=\rho\frac{\partial^2}{\partial t^2}(u+u_0), \tag{1}$$

where σ_x is the uniaxial stress in the bar in the longitudinal (x) direction. In this case, Hooke's law states that $\sigma_x=E^*\varepsilon_x$, where E^* is the complex Young's modulus and ε_x is the axial strain experienced by the material, defined by $\varepsilon_x\equiv\partial u/\partial x$. Substituting these two relations into Eq. (1) gives

$$E^*\frac{\partial^2u}{\partial x^2}=\rho\frac{\partial^2}{\partial t^2}(u+u_0), \tag{2}$$

which, in turn, leads to the following ordinary differential equation:

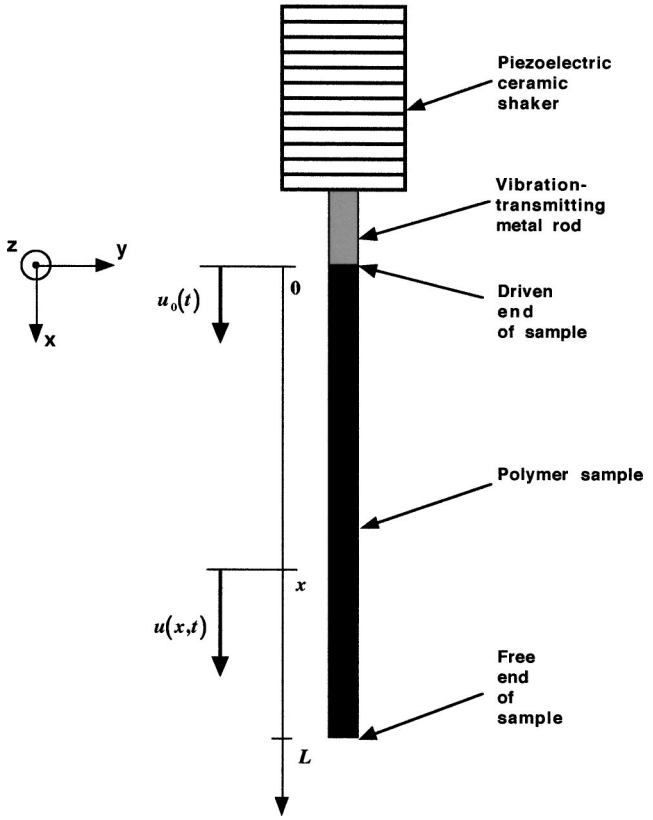


FIG. 1. Forced vibrations of a polymer sample in the form of a bar of length L with uniform cross section. The ceramic shaker and the vibration-transmitting metal rod correspond to the particular experimental setup used in this work. The sample excitation is measured at the bottom surface of the metal rod.

$$\frac{d^2U}{dx^2}+k^2U=-k^2U_0, \tag{3}$$

where k is the complex wave number, defined as $k\equiv\omega(\rho/E^*)^{1/2}=\omega/c_b^*$, and $c_b^*=(E^*/\rho)^{1/2}$ is the complex bar wave speed. The boundary conditions for the problem are zero relative displacement at $x=0$, i.e., $u(0,t)=0$ or $U(0)=0$, and zero stress at $x=L$, i.e., $\sigma_x(L,t)=0$ or $dU/dx_{\text{at } x=L}=0$, which lead to the following solution for Eq. (3):

$$U(x)=U_0[\cos(kx)+\tan(kL)\sin(kx)-1]. \tag{4}$$

Now that the equation of motion has been solved, the resonances of the sample (bar) can be studied. To do this, let us consider the complex ratio of the free-end displacement to the driven end displacement, Q^*, which, from Eq. (4), can be expressed as

$$Q^*=\frac{U(L)+U_0}{U_0}=\frac{1}{\cos(kL)}. \tag{5}$$

On the other hand, using complex notation, Young's modulus can be written as

$$E^*=E'+iE''=E[\cos(\delta)+i\sin(\delta)], \tag{6}$$

where E' is the elastic (or storage) modulus, E'' is the loss (or viscous) modulus, E is the magnitude, and $\tan(\delta)$ is the loss factor. Using these relations, Eq. (5) can then be expressed as

$$\frac{1}{Q^*} = \cos(kL) = \cos\left[\omega L\left(\frac{\rho}{E[\cos(\delta)+i\sin(\delta)]}\right)^{1/2}\right], \quad (7)$$

which can be separated into real and imaginary parts in the following manner:

$$\frac{1}{Q^*} = \cos\left[\omega L\left(\frac{\rho}{E}\right)^{1/2}\cos\left(\frac{\delta}{2}\right)\right]\cosh\left[\omega L\left(\frac{\rho}{E}\right)^{1/2}\sin\left(\frac{\delta}{2}\right)\right] + i\sin\left[\omega L\left(\frac{\rho}{E}\right)^{1/2}\cos\left(\frac{\delta}{2}\right)\right]\sinh\left[\omega L\left(\frac{\rho}{E}\right)^{1/2}\sin\left(\frac{\delta}{2}\right)\right]. \quad (8)$$

If the complex ratio is written as $Q^* = Qe^{i\theta}$, then at resonance, when $\theta = (2n-3)(\pi/2)$ and $n = 1,2,3,...$ is the resonance number, the left-hand side of Eq. (8) is

$$\frac{1}{Q^*} = i\frac{(-1)^{n+1}}{Q} \quad (n = 1,2,3,...). \quad (9)$$

Combining Eqs. (8) and (9) leads to the following system of two equations in two unknowns (E and δ):

$$\cos\left[\omega L\left(\frac{\rho}{E}\right)^{1/2}\cos\left(\frac{\delta}{2}\right)\right]\cosh\left[\omega L\left(\frac{\rho}{E}\right)^{1/2}\sin\left(\frac{\delta}{2}\right)\right] = 0, \quad (10)$$

$$\sin\left[\omega L\left(\frac{\rho}{E}\right)^{1/2}\cos\left(\frac{\delta}{2}\right)\right]\sinh\left[\omega L\left(\frac{\rho}{E}\right)^{1/2}\sin\left(\frac{\delta}{2}\right)\right] = \frac{(-1)^{n+1}}{Q}.$$

One can easily show that the solutions to Eq. (10) are

$$\delta = 2\tan^{-1}\left[\frac{\sinh^{-1}(1/Q)}{(2n-1)\frac{\pi}{2}}\right], \quad (11)$$

and

$$E = \rho\left[\frac{4Lf_{\text{res}}}{(2n-1)}\cos\left(\frac{\delta}{2}\right)\right]^2. \quad (12)$$

One should again emphasize that Eqs. (11) and (12) are valid at resonance only, that f_{res} is the resonance frequency, and that Q is the amplitude of the displacement ratio. These equations are in agreement with the theory presented in Ref. 2, for the case of an end mass equal to zero. One can see from the above analysis that measuring f_{res} and Q at resonance, and using Eqs. (11) and (12) provides a direct method for determining the complex Young's modulus of a sample with no mass attached to its free end. Once the modulus amplitude and the loss factor are known, the elastic and loss moduli can be computed using

$$E' = E\cos(\delta) = \rho\left[\frac{4Lf_{\text{res}}}{(2n-1)}\cos\left(\frac{\delta}{2}\right)\right]^2\cos\left\{2\tan^{-1}\left[\frac{\sinh^{-1}(1/Q)}{(2n-1)\frac{\pi}{2}}\right]\right\},$$

$$E'' = E\sin(\delta) = \rho\left[\frac{4Lf_{\text{res}}}{(2n-1)}\cos\left(\frac{\delta}{2}\right)\right]^2\sin\left\{2\tan^{-1}\left[\frac{\sinh^{-1}(1/Q)}{(2n-1)\frac{\pi}{2}}\right]\right\}. \quad (13)$$

B. Wave-speed measurements

Let us again consider the bar shown in Fig. 1. Instead of a continuous harmonic signal, the shaker is excited with a short burst consisting of a gated sinusoidal signal. In this case, monitoring the propagation of this burst allows one to measure both the wave speed and the attenuation in the material. The wave-speed magnitude, c_b, is obtained in a straightforward manner by measuring the time necessary for the burst to travel a distance d along the sample. The signal attenuation (due to the material's internal damping) is obtained from the change in signal amplitude over the same distance. Assuming that the displacement perturbation corresponding to the traveling burst can be written as

$$u(x,t) = U_0 e^{i(\omega t - kx)}, \quad \text{for} \quad t_1 \leqslant t \leqslant t_2, \quad (14)$$

and recalling from the previous section that the complex wave number k is given by

$$k = \omega\left(\frac{\rho}{E^*}\right)^{1/2}, \quad (15)$$

then, using Eq. (6), one obtains

$$k = \omega\left(\frac{\rho}{E[\cos(\delta)+i\sin(\delta)]}\right)^{1/2} = \omega\left(\frac{\rho}{E}\right)^{1/2}\left[\cos\left(\frac{\delta}{2}\right) - i\sin\left(\frac{\delta}{2}\right)\right] = k' - i\alpha, \quad (16)$$

and

$$u(x,t) = U_0 e^{-\alpha x}\cdot e^{i(\omega t - k' x)}, \quad (17)$$

where α is the attenuation coefficient. Thus, as can be seen from Eq. (17), α can be computed from the amplitude of the displacement signal measured at two locations spaced apart by a distance d, according to

$$\alpha = \frac{1}{d}\ln\left(\left|\frac{u(x,t)}{u(x+d,t)}\right|\right). \quad (18)$$

The magnitude of Young's modulus and its loss factor $\tan(\delta)$ can then be obtained from

$$E = \rho c_b^2, \quad (19)$$

and, from Eq. (16)

$$\delta = 2\sin^{-1}\left(\frac{\alpha c_b}{\omega}\right). \tag{20}$$

Finally, the elastic and loss moduli are given by

$$\begin{aligned} E' &= E\cos(\delta) = \rho c_b^2 \cos\left[2\sin^{-1}\left(\frac{\alpha c_b}{\omega}\right)\right], \\ E'' &= E\sin(\delta) = \rho c_b^2 \sin\left[2\sin^{-1}\left(\frac{\alpha c_b}{\omega}\right)\right]. \end{aligned} \tag{21}$$

C. Dispersive effects

It has been assumed so far that the effects of lateral inertia in the sample were negligible. The purpose of this section is to reexamine this assumption and to determine under which conditions these effects need to be taken into account in the measurement of Young's modulus. Let us again consider the sample in the form of a bar depicted in Fig. 1. Taking into account the lateral displacements resulting from the Poisson coupling of the longitudinal wave, and keeping the assumption of uniaxial stress in the bar, the general Hooke's law can be written as

$$\varepsilon_x = \frac{1}{E^*}\sigma_x, \quad \text{and} \quad \varepsilon_y = \varepsilon_z = -\frac{\eta}{E^*}\sigma_x, \tag{22}$$

where η is the material's Poisson's ratio and the subscripts denote the spatial directions (see Fig. 1). If $u(x,t)$, $v(x,t)$, and $w(x,t)$ denote the displacements in the x, y, and z directions, respectively, then Eq. (22) can be used to compute the lateral displacements in terms of the longitudinal displacement as

$$v = -\eta y\frac{\partial u}{\partial x} \quad \text{and} \quad w = -\eta z\frac{\partial u}{\partial x}, \tag{23}$$

where y and z are the coordinates of a point in the sample cross section (the center of the coordinate system coincides with the center of the cross section). Using Eq. (23), it is possible to formulate the potential and kinetic energies of the bar, and to apply Hamilton's principle to derive the equation of motion corresponding to longitudinal vibrations with lateral inertia effects, known as the Love equation.[11] The details of this derivation are somewhat lengthy and can be found in Ref. 12. Love's equation of motion is

$$E^*\frac{\partial^2 u}{\partial x^2} + \rho\eta^2 K^2\frac{\partial^4 u}{\partial x^2 \partial t^2} = \rho\frac{\partial^2 u}{\partial t^2}, \tag{24}$$

where K^2 is the polar radius of gyration of the cross section. Considering a solution to Eq. (24) of the form

$$u(x,t) = Ue^{i\omega(t-[x/c^*])}, \tag{25}$$

where c^* is the velocity of the longitudinal waves, and substituting (25) into (24), yields

$$-E^*\frac{\omega^2}{c^{*2}} + \rho\eta^2 K^2\frac{\omega^4}{c^{*2}} = -\rho\omega^2. \tag{26}$$

Equation (26) can be rearranged to yield

$$\left(\frac{E^*}{\rho}\right)^{1/2} = c_b^* = c^*\left[1+\left(\frac{\eta K\omega}{c^*}\right)^2\right]^{1/2}. \tag{27}$$

Equation (27) indicates that elastic wave propagation in the bar is dispersive and relates the complex bar speed c_b^* to the complex longitudinal wave speed c^* at a given frequency, ω. The same result could have been obtained from the exact solution, derived by Pochhammer, to the wave propagation problem in a circular rod of radius a, in the low-frequency approximation[13] (i.e., when the radius a is much smaller than the wavelength, and when, consequently, the uniaxial stress assumption holds). This was also observed by Rayleigh,[14] who derived Eq. (27) in the case of a circular rod. For a square cross section of side length s, as is the case in our measurements, $K^2 = s^2/6$, and Eq. (27) becomes

$$\left(\frac{E^*}{\rho}\right)^{1/2} = c_b^* = c^*\left[1+\frac{1}{6}\left(\frac{\eta s\omega}{c^*}\right)^2\right]^{1/2}. \tag{28}$$

As can be seen from Eq. (28), the lateral inertia effects increase with increasing values of Poisson's ratio, frequency, and sample cross section; they are also more substantial at lower wave speeds. Therefore, when the experimental conditions are such that these effects are significant, the measured wave speed c^* differs from the bar wave speed c_b^*, and Eq. (28) must be used to compute Young's modulus. Practically, it is assumed that the correction factor (i.e., the second term inside the brackets) is small enough so that Eq. (28) applies to the *magnitude* of the complex speeds. The experimentalist needs to be especially cautious when measuring Young's modulus on materials with large values of Poisson's ratio, at high frequency. In such cases, or in cases where Poisson's ratio is not known, it is advisable to minimize the lateral dimensions of the sample, in order to minimize the dispersive effects.

III. SAMPLE PREPARATION AND EXPERIMENTAL APPARATUS DESCRIPTION

A. Sample preparation

Samples are cut in the shape of a bar with a square cross-sectional area. The maximum sample length is limited by the pressure vessel dimensions and is about 27 cm. As explained above, it is desirable to use a sample with a minimal cross section. In the case of Rubatex® R451N, the material comes in the form of 6.35-nm (0.25-in.)-thick sheets that were cut in 0.25-in.-wide strips using a mat cutter. The Goodrich Thorodin™ AQ21 sample used in this study was cut into a bar with a 5- by 5-mm cross-sectional area. The top and bottom ends of each sample were then cut with a razor blade in order to get plane and smooth surfaces, which are necessary for a good bond between the sample and the metal of the excitation-transmitting rod shown in Fig. 1. The end surfaces also need to be perpendicular to the axis of the sample so that the latter can be mounted in a perfectly vertical position inside the apparatus, and therefore be excited in a primarily longitudinal mode, minimizing spurious flexural motions.

For materials whose color and texture make them good light scatterers, no further preparation is needed. However,

other materials, such as Rubatex® R451N (which is black and has a matte finish), and Goodrich Thorodin™ AQ21 (which is yellow and somewhat translucent), do not scatter sufficient amounts of light and therefore require a surface treatment in order to be measured with a laser vibrometer. An obvious requirement is that the surface treatment not change the elastic properties of the sample. With this in mind, the following treatments have been used. In the case of Rubatex® R451N, for the measurement of the motion of the bottom end of the sample, a small piece of metallized tape (about the same dimensions as the cross-sectional area) is glued (using a cyanoacrylate adhesive) to the end of the sample. This piece of tape does not constrain the longitudinal motion, and its mass is negligible with respect to that of the sample. For measurement of the motion of the lateral sides of the sample (resulting from the Poisson-coupled longitudinal motion), a layer of talc powder is deposited on these surfaces. The load resulting from the presence of the talc is negligible and its powdery nature allows the sample to move in the lateral direction without any constraint. In the case of Goodrich Thorodin™ AQ21, a thin layer of white correction fluid (''white-out'') was applied at the bottom end of the sample. The same treatment as for the Rubatex® R451N material (metallized tape) could have been used, but the motivation for trying the correction fluid was to compare its performance to that of the tape. The white fluid was found to be a slightly better light scatterer than the tape; however, both are perfectly adequate for the resonance measurements. For the side measurements on Goodrich Thorodin™ AQ21, a thin layer of gold-colored ink was used, which proved to be a better reflector than the talc powder. The absence of adverse effects from the tape, the powder, and the ink on the sample motion was verified experimentally by measuring the vibrations of samples with and without these surface treatments: the measured displacements were identical in both cases.

As a final note specific to Rubatex® R451N, before cutting a sample, the actual first preparatory step consisted of following the manufacturer's recommendation of placing the sheet of foam in an oven, set at 71 °C (160 °F), for 24 h, in order to completely cure the material and insure stable properties. It was found, however, that even after 4 days of this treatment, measured properties had a tendency to change with time. Therefore, the data that are presented below should be considered a ''snapshot'' of the Rubatex® R451N material constants.

B. Experimental apparatus

The experimental apparatus is shown in Fig. 2. The sample is glued to a 2.5-cm-long vibration-transmitting metal rod (with approximately the same cross-sectional area as the sample), which is itself glued to a shaker, using a cyanoacrylate adhesive. The shaker is made of ten piezoelectric ceramic discs [EC64 from EDO Corporation: 1.78 cm (0.700-in.) outer diameter $\times$0.22-cm (0.085-in.)-thick] plus one depolarized and unelectroded disk at each each end of the stack, for shielding purposes; the disks are glued together using an epoxy resin. The shaker, in turn, is epoxied to a threaded mounting piece that attaches to the top of the pressure vessel. To prevent delaminations caused by temperature cycling, the vibration-transmitting rod and the mounting piece are made of Invar, which has a very low coefficient of thermal expansion, comparable to that of the ceramic used in the shaker.

The pressure vessel [inside dimensions: 38.72 cm (15.25 in.)L$\times$10.2 cm (4 in.)W$\times$7.6 cm (3 in.)D] is made of 303 stainless steel and it is designed to be used up to 3.45 MPa (500 psi). It features three 1.9-cm (0.75 in.)-thick glass windows for optical access to the sample: one round window [Ø 3.5 cm (1.375 in.)] located at the bottom of the vessel and two long windows [28.8 cm (11.35 in.)L $\times$3.4 cm (1.35 in.)W] located along its sides. They are sealed inside the vessel with a 1.59-mm (0.0625 in.)-thick Aramid gasket used in conjunction with a silicone RTV sealant. An RTD sensor (Sensor Scientific PT1000) is used to measure the temperature inside the vessel. Pressure is created with a compressed air cylinder connected to the vessel (not shown in Fig. 2) and both pressure and temperature are monitored using a Heise PM Indicator gauge.

Outside the pressure vessel, three laser Doppler vibrometer sensor heads (Polytec CLV-800-ff), capable of measuring out-of-plane (normal) vibrations, are mounted on positioning slides (two sets of 5-in. and 15-in. Velmex Bislides for the side lasers; two 25-mm and one 110-mm National Aperture Mini Stages for the bottom one). The entire system (vessel, lasers, and slides) is placed inside an environmental chamber (Thermotron S-27C) for temperature control, and needs to be operated from outside that chamber. For that purpose, the slides are remotely driven by two controllers (Velmex VP 9000 and National Aperture Servo 1000), which allow precise positioning of the vibrometers. Also, a miniature video camera connected to a television monitor is mounted next to each sensor head; it is focused on the area where the laser beam illuminates the surface on the sample, and provides a direct visual check of the beam position as well as a view of the sample from three different angles. A neutral density filter is placed in front of each camera in order to reduce the light intensity and to improve the laser spot resolution on the monitor. Two small flashlights are used to enhance the image of the sample. Because of the small aperture afforded by the bottom window and lack of space in that area of the setup, the corresponding camera is positioned at a right angle and a small mirror is employed to obtain a view of the bottom of the sample.

The shaker is excited by two types of signal, depending on which measurement is performed. For the resonance measurements, a continuous signal with varying frequency is used, provided by a lock-in amplifier (Stanford Research Systems model SR850). For the wave-speed measurements, a burst signal composed of six cycles at a given frequency is used, provided by a function generator (Wavetek model 80). In both cases, signals are amplified by a power amplifier (Krohn-Hite model 7500) before being sent to the shaker. Surface motion signals measured by the sensor heads are electronically processed by the vibrometer controller: CLV-1000 (bottom laser) or CLV-2000 (side lasers). These signals are filtered (Krohn-Hite filter model 3382) before being displayed on an oscilloscope (Tektronix model TDS 3014). In

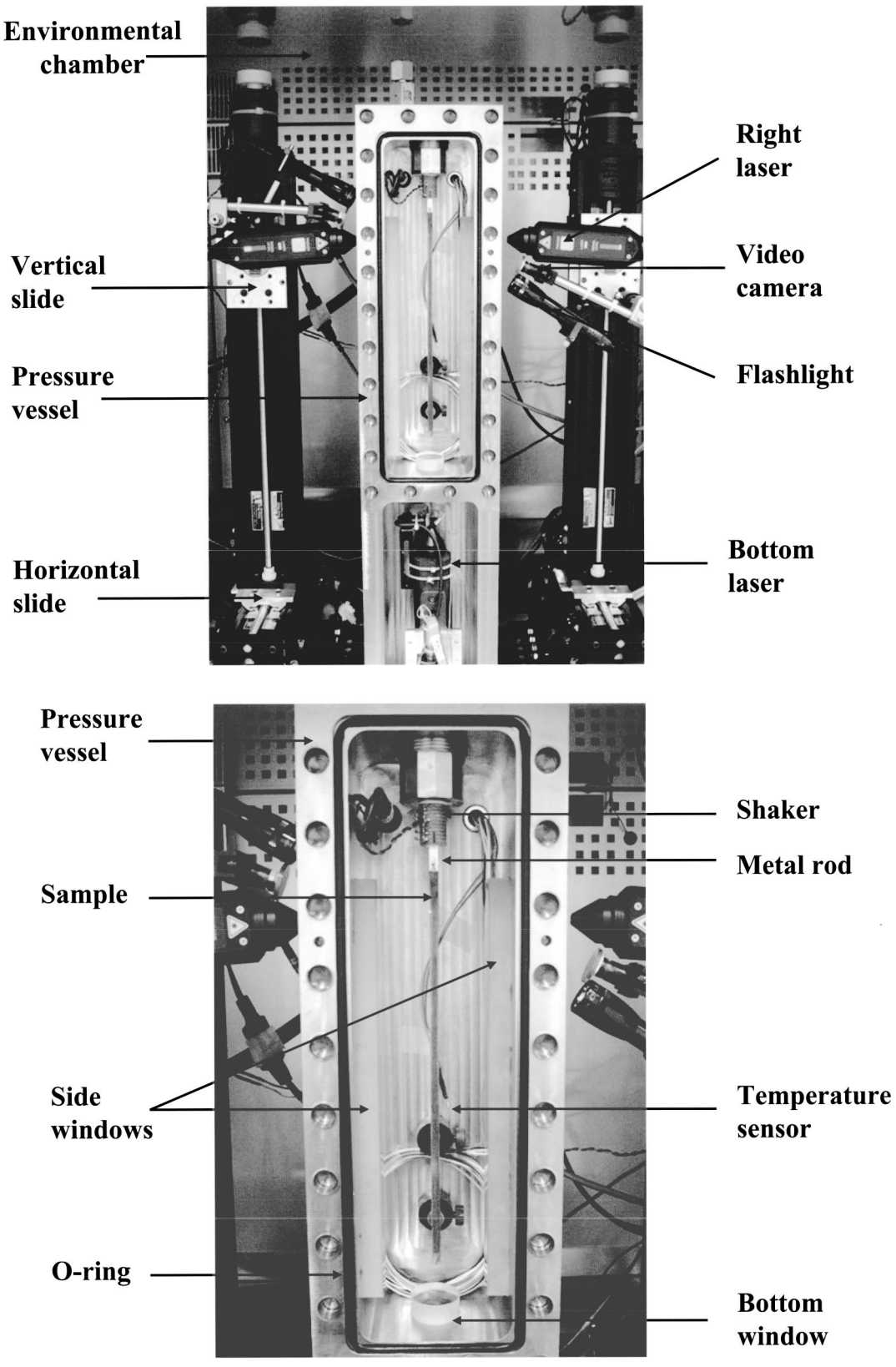


FIG. 2. Picture of the experimental setup. (a) General view. (b) Close-up.

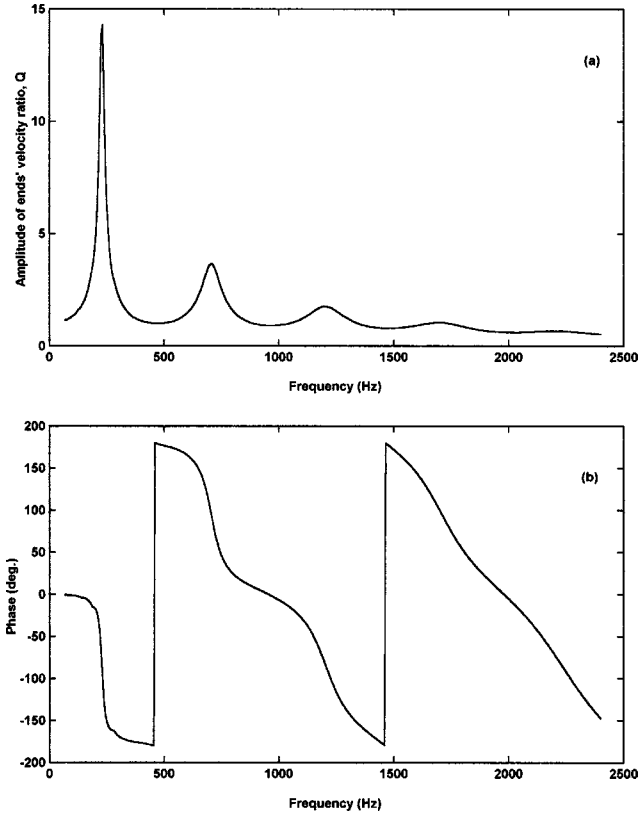


FIG. 3. Resonance plots obtained on a Rubatex® R451N sample at 30 °C and ambient pressure, displaying four measurable resonances. (a) Amplitude plot of the ratio of the velocity of the free end to that of the driven end. (b) Corresponding phase plots.

the case of the resonance measurements, the velocity signals are also input to the lock-in amplifier. Both the oscilloscope and the lock-in amplifier are equipped with a GPIB bus utilized to transfer velocity signals to a notebook computer for data processing.

IV. MEASUREMENT PROCEDURE AND DATA PROCESSING

A. Resonance measurements

Based on the theory described in Sec. II A, the procedure to implement the resonance measurement technique is as follows. These measurements are obtained using the bottom laser only. The lock-in amplifier, in the frequency sweep mode, excites the shaker with a continuous sine wave whose frequency ranges from a lower limit to an upper limit. The amplifier also records the amplitude and phase of the signals output by the laser vibrometer, the latter being employed to measure both the velocity of the driven end (the metal rod without the sample adhered to it) and the velocity of the free end. These signals, digitized with 401 points by the lock-in amplifier, are combined numerically in MATLAB to produce amplitude and phase plots of the sample frequency response. First, a ''global'' sweep is performed over an extended frequency range in order to identify all the measurable resonances of the sample. The resulting plots are shown in Fig. 3 over a 50–2400-Hz range, which was the typical range used for the measurements on the Rubatex® R451N sample. Each resonance frequency is approximately determined from the phase plot at points where the phase ϕ is equal to $(2n-3)\times 90$. A second sweep is then performed over a narrow frequency range centered around each of these frequencies, and the resulting plots are used to obtain more accurate values of f and Q at resonance. The narrow range has a span of 100 Hz, which, combined with the 401 points of digitization, allows for the determination of the resonance frequency with a resolution of 0.25 Hz. These data, along with the length and density of the sample, are then input into Eq. (13) to compute the elastic and loss moduli. The maximum measurement errors associated with these two quantities can be estimated as follows. The 0.25-Hz resolution results in a frequency uncertainty of 0.5% at 50 Hz; assuming a ± 0.5-mm error on the sample length measurement results in a 0.3% length uncertainty for a 15-cm-long sample (the smallest sample used in the system). The systematic errors from the laser vibrometer decoder module and from the lock-in amplifier do not affect the measurement of Q, the amplitude ratio, because it is a relative measurement. Therefore, these errors are not taken into account. This analysis results in a total uncertainty of 1.6% for both the elastic and loss moduli. One should note that this does not include the uncertainty related to the density, which affects all measurements (elastic and loss moduli, from both resonance and wave-speed techniques) in an identical manner.

B. Wave-speed measurements

In the method described above, internal damping in materials always limits the number of identifiable resonances, and, therefore, sets an upper frequency limit on the method for a given sample. One way to extend measurements to higher frequency regions is to reduce the length of the sample, which, as Eq. (12) reveals, would shift the value of resonance frequencies higher. In theory, successive cuts can be made to produce increasingly shorter samples to cover the frequency range of interest, as long as the sample maintains an appropriate dimension ratio (i.e., $L \gg$ lateral dimensions). However, this practice possesses two disadvantages. First, if the sample under study is inhomogeneous, then cutting the sample into smaller lengths effectively results in measuring a different sample after each cut, introducing additional measurement errors. Second, each sample shortening requires opening the pressure vessel and the environmental chamber, and the subsequent readjusting of the necessary pressure and temperature conditions, rendering the process quite time consuming. A better way to extend the experimental Young's modulus frequency range is to measure the wave speed and the attenuation of longitudinal waves propagating inside a sample. In our apparatus, this technique is implemented with the two side lasers measuring the lateral component of a longitudinal excitation propagating along the sample. A short burst is used to excite the shaker, which excites a longitudinal wave in the sample and the resulting lateral motion (due to the Poisson's ratio effect) is then recorded at two different locations along the length of the sample. Two laser vibrometers located on opposite sides of the sample are needed in order to insure that only the lateral motion (which is a symmetrical motion) produced by the longitudinal wave is measured. Any other spurious vibrations (due to mounting im-

perfections or outside vibrations) that the sample may experience create bending motions, which manifest themselves in the form of antisymmetrical displacements. Therefore, these spurious vibrations can be systematically eliminated by adding the signals from each vibrometer.

The amount of attenuation (which, above the glass transition temperature, increases with frequency) in the material sets both the lower- and the upper-frequency limits of the wave-speed method. Indeed, it is imperative that the laser vibrometers measure only the direct signal produced by the shaker without any superimposed reflections from the ends of the sample. Thus, sufficient attenuation needs to be present in order to eliminate these reflections before they reach the measurement locations; this determines the lower-frequency limit (for a given sample length). Conversely, the upper limit is attained when attenuation is so large as to prevent accurate measurement of the signal amplitude at the location furthest away from the shaker. Thus, the actual frequency range of the measurements described in this article depends only on the attenuation of the particular material under study, and will therefore vary from one material to the other. For the Rubatex® R451N sample, combining the resonance and the wave-speed techniques, the range was found to be approximately 200–8000 Hz. These values, of course, like the attenuation, also depend on temperature and pressure.

Practically, the wave-speed technique is implemented as follows. The function generator excites the shaker with a burst signal composed of 6 cycles at a given frequency, with a 1-s repetition rate. For example, frequencies ranging from 4500 to 8000 Hz in steps of 500 Hz were used for the Rubatex® R451N sample. Both laser sensor heads are positioned at the same height, about 2 cm below the shaker; this constitutes the top location. At each frequency, signals from each vibrometer are displayed on the oscilloscope, which digitizes them using 10 000 points. Sixty-four averages are taken, and the signals from each side are added; the resulting signal is transferred to a computer for further processing. This is repeated at the distant location, 10 cm below the top one. Typical signals measured at both locations and used for wave speed and attenuation computation are shown in Fig. 4. In MATLAB, signals are windowed and zero padded before being subjected to a fast Fourier transform. The center frequency of each Fourier-transformed signal is determined, and, at that frequency, the amplitude A and the phase φ of the ratio of the top location signal to that of the distant location are computed. These yield the attenuation and the wave speed, respectively, according to

$$\alpha = \frac{\ln(A)}{d}, \tag{29}$$

$$c = \left(\frac{\omega}{\varphi}\right) d. \tag{30}$$

These data, along with the density of the sample, are then input into Eq. (21) to compute the elastic and loss moduli. The uncertainties associated with these measurements are assessed as follows. Assuming a ± 0.5-mm error on distance measurement results in a 0.5% uncertainty for d; the systematic errors associated with the vibrometers and the oscilloscope do not affect the measurement of the ratio A; this results in a maximum total error on the order of 1.0% for the elastic modulus and on the order of 2.0% for the loss modulus. As previously mentioned, these values do not include the error associated with density measurements.

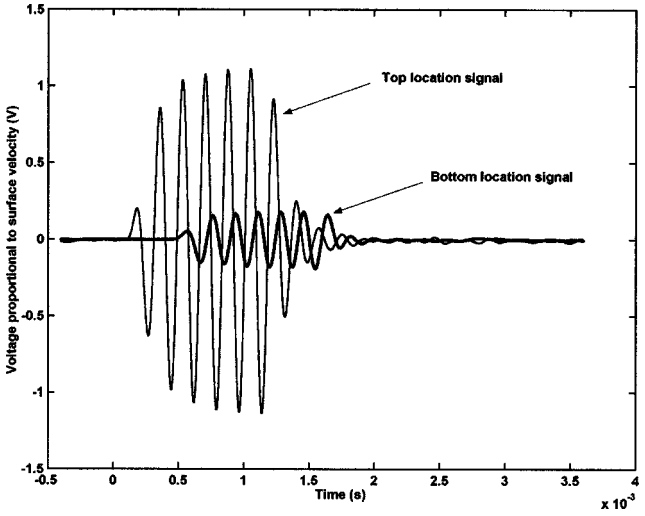


FIG. 4. Velocity signals measured at 5500 Hz by the side laser vibrometers at two locations spaced 10 cm apart on the surface of a Rubatex® R451N sample (30 °C and ambient pressure). The relative phase and amplitude of these signals are used to compute, respectively, the speed and attenuation of longitudinal waves propagating inside the sample.

V. EXPERIMENTAL RESULTS

A. Rubatex® R451N

The following data were obtained on a Rubatex® R451N neoprene sample, with a density of 579 kg/m3, and cut to a length of 23.3 cm. These two values were measured at room temperature and ambient pressure. The subsequent dimensional changes resulting from varying the hydrostatic pressure and the temperature were assessed using one of the side lasers and the positioning slide controller: the laser beam was focused at the bottom of the sample whose change of position was recorded for the various pressures and temperatures used. This yielded the length of the sample, whose value was used to estimate the density of the material. One should note that, for samples that have been measured in the bulk modulus system of the Acoustic Material Laboratory, the dependence of density on pressure and temperature is known. However, Rubatex® R451N is too soft to be measured in that system. The data shown below are not corrected for dispersive effects, as Poisson's ratio of Rubatex® R451N (which is on the order of 0.25) is small enough to render the corrections negligible.

1. Young's modulus as a function of hydrostatic pressure

Figures 5 and 6 display resonance and wave-speed measurement results in the form of the real and imaginary parts of the modulus, as a function of pressure. Data were obtained at 30 °C, and 0 (i.e., ambient), 69 and 138 kPa (0, 10, and 20 psi). At least 30 min were allowed to elapse after each new pressure setting, and the pressure was always increased.

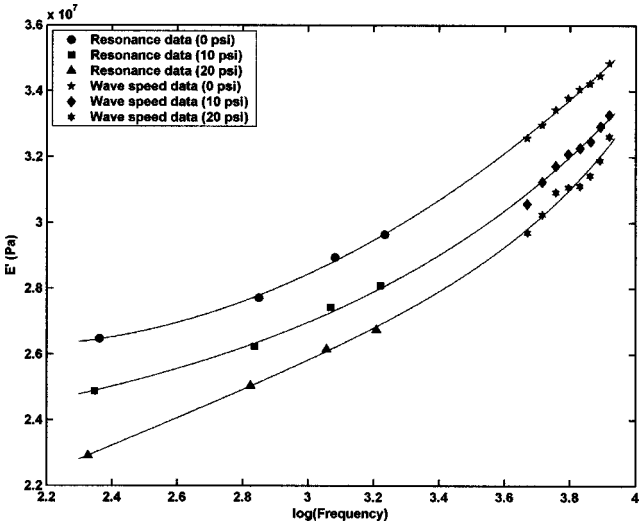


FIG. 5. Elastic Young's modulus of Rubatex® R451N measured at 30 °C, as a function of hydrostatic pressure. The solid lines are curve fits.

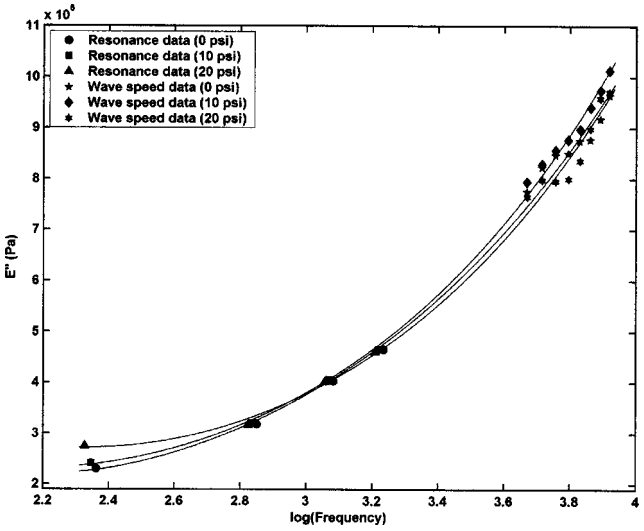


FIG. 6. Loss Young's modulus of Rubatex® R451N measured at 30 °C, as a function of hydrostatic pressure. The solid lines are curve fits.

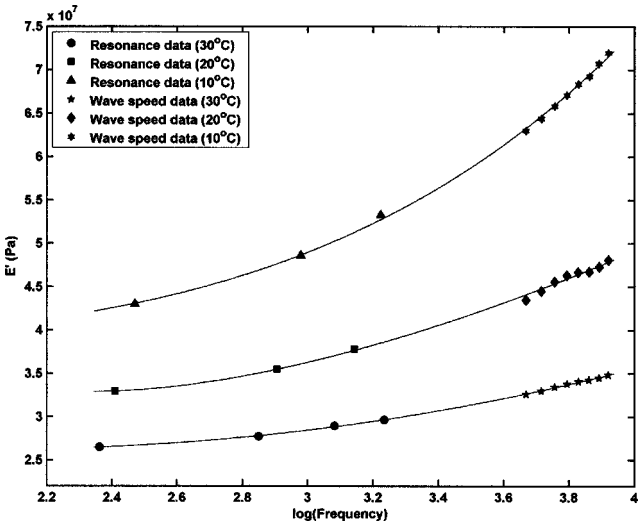


FIG. 7. Elastic Young's modulus of Rubatex® R451N measured at ambient pressure, as a function of temperature. The solid lines are curve fits.

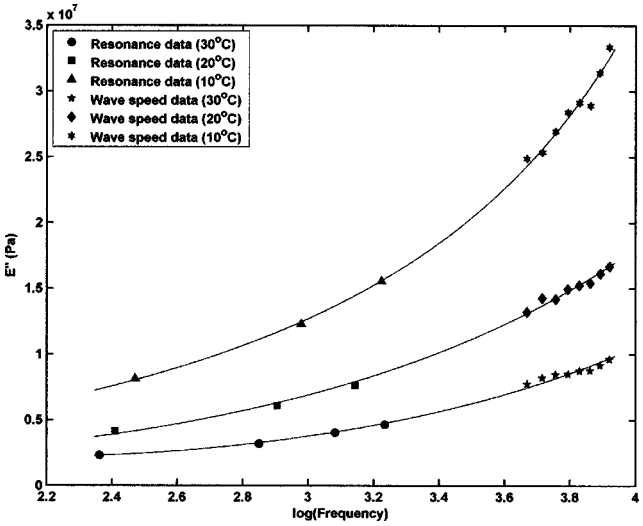


FIG. 8. Loss Young's modulus of Rubatex® R451N measured at ambient pressure, as a function of temperature. The solid lines are curve fits.

These plots clearly show that the two techniques provide results that are compatible with each other, and that the wave-speed approach allows one to substantially extend the measurement frequency range beyond what the classical resonant bar method can offer, without having to cut the sample.

Figure 5 indicates that the material gets softer with increasing hydrostatic pressure, as the walls of the closed cells buckle and become more compliant. The loss modulus, however, is relatively unaffected by pressure changes, as shown in Fig. 6.

2. Young's modulus as a function of temperature

Figures 7 and 8 display resonance and wave-speed measurement results in the form of the real and imaginary parts of the modulus, respectively, as a function of temperature.

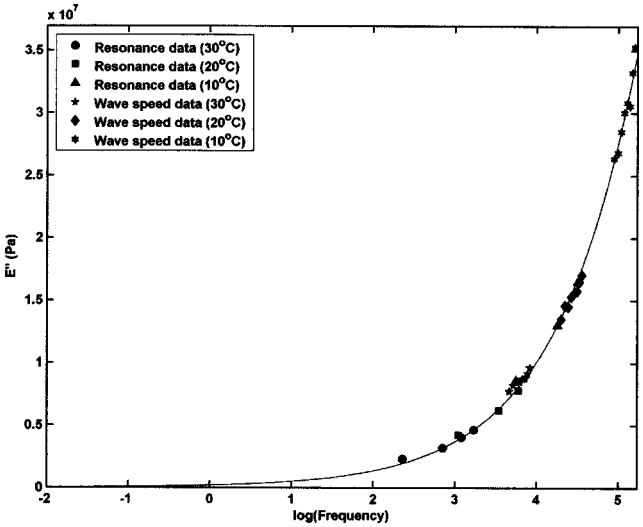


FIG. 9. Loss Young's modulus of Rubatex® R451N at ambient pressure and at 30 °C, resulting from horizontal shifts applied to the 20 °C and 10 °C data sets. The 20 °C and 10 °C data have been corrected by multiplicative factors of $T_{30}\rho_{30}/T_{20}\rho_{20}$ and $T_{30}\rho_{30}/T_{10}\rho_{10}$, respectively. The solid line is a curve fit.

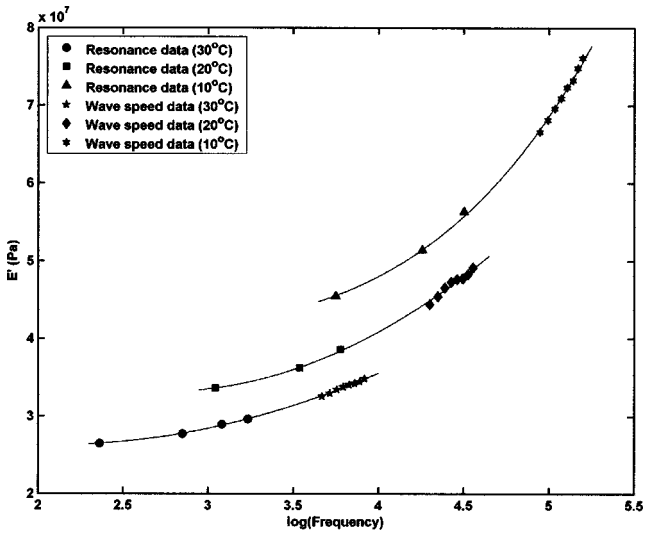


FIG. 10. Elastic Young's modulus of Rubatex® R451N at ambient pressure and at 30 °C, resulting from horizontal shifts applied to the 20 °C and 10 °C data sets. The shift factors are the same as the ones used for the loss modulus. The 20 °C and 10 °C data have been corrected by multiplicative factors of $T_{30}\rho_{30}/T_{20}\rho_{20}$ and $T_{30}\rho_{30}/T_{10}\rho_{10}$, respectively. The solid lines are curve fits.

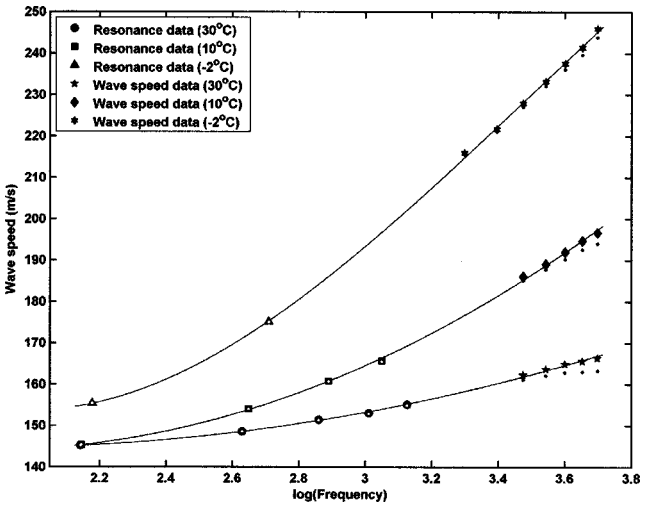


FIG. 12. Wave speed of Goodrich Thorodin™ AQ21 measured at 50 psi, as a function of temperature. The dots represent wave-speed values uncorrected for dispersion. In the case of the resonance data, the correction is negligible, and the white dots (uncorrected values) appear in the center of the black-filled markers (corrected values). In the case of the wave-speed data, the correction is more significant, and the black dots are distinct from the black-filled markers.

Data were obtained at ambient pressure, and at 30 °C, 20 °C, and 10 °C. The sample was allowed to equilibrate overnight between each temperature. Figures 7 and 8 indicate that both the elastic and the loss moduli increase significantly with decreasing temperature, as expected above the glass transition temperature. Only three resonances were measurable at the two lowest temperatures, due to the increased loss.

Finally, the time–temperature superposition principle can be applied to these data in the following manner. First, the loss modulus temperature curves are shifted horizontally to get the best possible fit to a smooth curve, as illustrated in Fig. 9. In that figure, the reference temperature is 30 °C, and the horizontal shift factors applied to the 20 °C and 10 °C data sets are $a_{20}=4.3$ and $a_{10}=19.0$, respectively. These last two data sets have been corrected by multiplicative factors of $T_{30}\rho_{30}/T_{20}\rho_{20}$ and $T_{30}\rho_{30}/T_{10}\rho_{10}$, respectively (where the Ts are temperatures (in Kelvin) and the ρ's are the corresponding densities), as specified by Ferry.[15] The shift factors are then used to shift the elastic modulus horizontally by the same amount, producing the curves shown in Fig. 10. In that figure, the 20 °C and 10 °C data have also been corrected by multiplicative factors of $T_{30}\rho_{30}/T_{20}\rho_{20}$ and $T_{30}\rho_{30}/T_{10}\rho_{10}$, respectively. Finally, a vertical shift is applied to the temperature segments of Fig. 10, in order to obtain the smooth curve displayed in Fig. 11. The additive vertical shift factors are $\nu_{20}=-5.0\times10^6$ Pa and $\nu_{10}=-12.4\times10^6$ Pa for the 20 °C and 10 °C data sets, respectively. The physical meaning of these vertical shifts may be related to the static modulus dependence on temperature. However, in the case of the Rubatex® R451N sample, their significance is not entirely clear, as the elastic modulus of this material has been ob-

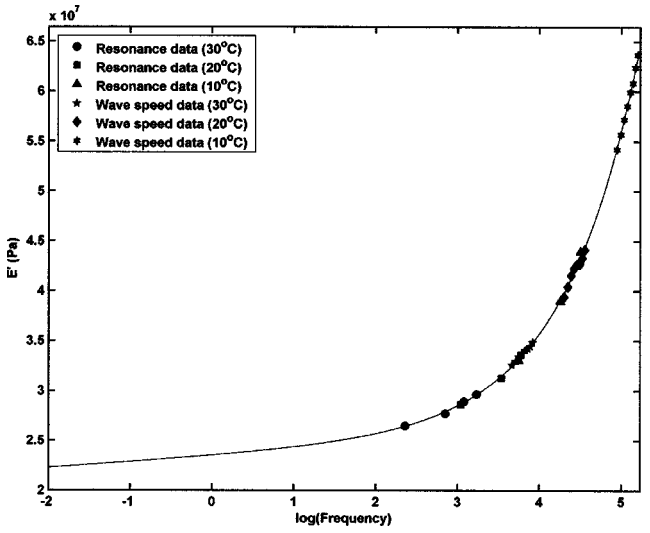


FIG. 11. Elastic Young's modulus of Rubatex® R451N at ambient pressure and at 30 °C, resulting from both horizontal and vertical shifts applied to the 20 °C and 10 °C data sets. The solid line is a curve fit.

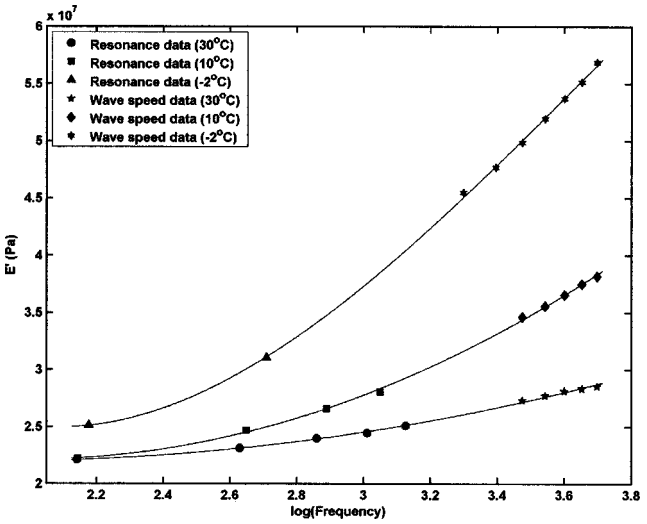


FIG. 13. Elastic Young's modulus of Goodrich Thorodin™ AQ21 measured at 50 psi, as a function of temperature. The solid lines are curve fits.

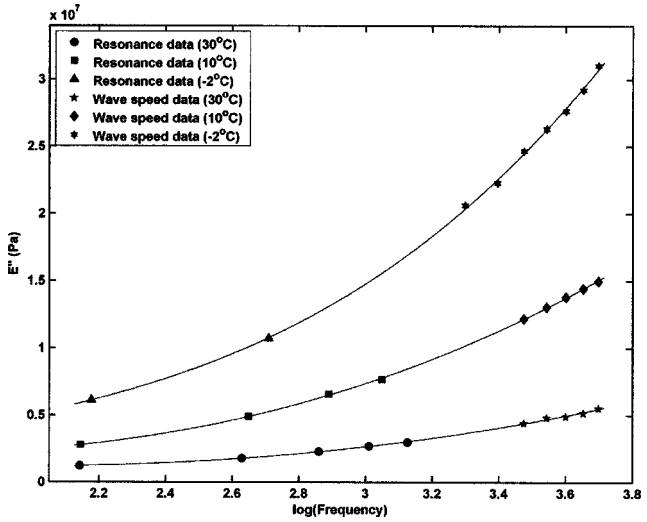


FIG. 14. Loss Young's modulus of Goodrich Thorodin™ AQ21 measured at 50 psi, as a function of temperature. The solid lines are curve fits.

served to change with time (by repeating measurements on subsequent days), and the data used to generate Fig. 11 were collected over a period of 3 days (one temperature per day).

One should note that some authors[16] use the loss factor instead of the loss modulus to obtain the horizontal shift factors. Other authors[6] even perform the horizontal shift first on the modulus magnitude and use the resulting factors to shift the loss factor data (their reasoning is that the magnitude is more accurately measured than the loss factor). It has been found, for the 13 samples of various materials measured to date in our system, that the best results are obtained using the approach described in the preceding paragraph. The reason for choosing the loss modulus to determine the initial horizontal shift is that it does not require any vertical correction: indeed, it can be observed in Fig. 9 (where only a horizontal shift has been applied), that the curve fit to the loss modulus data tends to a value of zero in the zero frequency limit (i.e., the dc limit), as must be the case for any viscoelastic material. All samples measured to date have required the adjustment of *two* parameters for the elastic modulus, the horizontal and the vertical shift factors. This being the case, the absence of *a priori* knowledge of the horizontal factors would result in nonunique curves, each corresponding to a different combination of horizontal and vertical shifts.

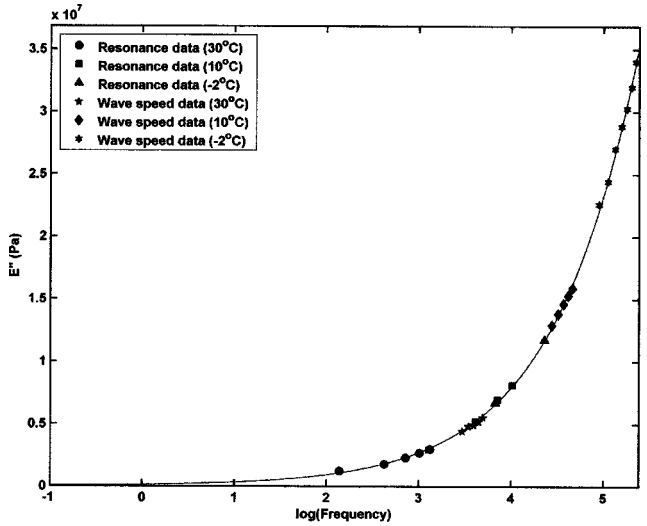


FIG. 15. Loss Young's modulus of Goodrich Thorodin™ AQ21 at 50 psi and at 30 °C, resulting from horizontal shifts applied to the 10 °C and −2 °C data sets. The 10 °C and −2 °C data have been corrected by multiplicative factors of $T_{30}\rho_{30}/T_{10}\rho_{10}$ and $T_{30}\rho_{30}/T_{-2}\rho_{-2}$, respectively. The solid line is a curve fit.

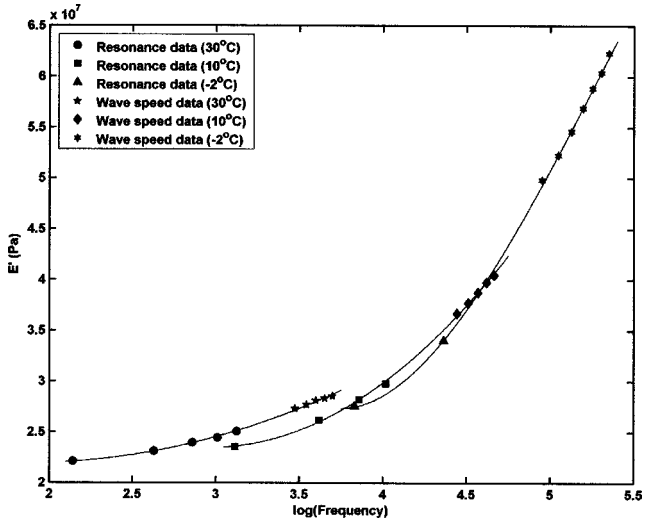


FIG. 16. Elastic Young's modulus of Goodrich Thorodin™ AQ21 at 50 psi and at 30 °C, resulting from horizontal shifts applied to the 10 °C and −2 °C data sets. The shift factors are the same as the ones used for the loss modulus. The 10 °C and −2 °C data have been corrected by multiplicative factors of $T_{30}\rho_{30}/T_{10}\rho_{10}$ and $T_{30}\rho_{30}/T_{-2}\rho_{-2}$, respectively. The solid lines are curve fits.

B. Goodrich Thorodin™ AQ21

The following data were obtained on a Goodrich Thorodin™ AQ21 sample, with a density of 1050 kg/m3, and cut to a length of 26.0 cm. This material is a solid, nearly incompressible polyurethane, with a Poisson's ratio value close to 0.5. Consequently, its Young's modulus does not exhibit any

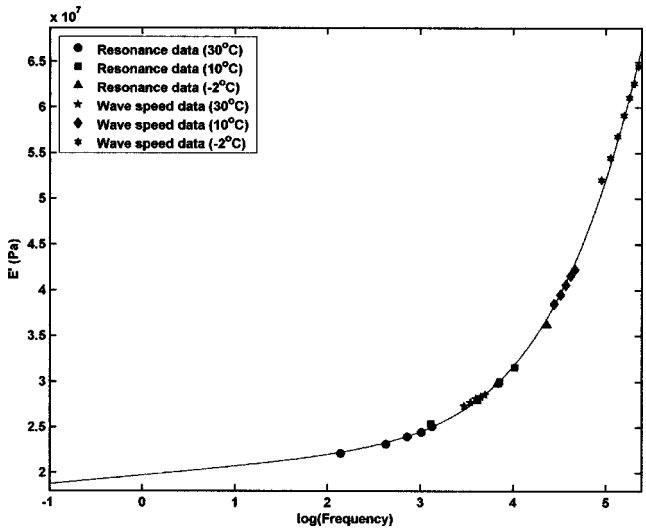


FIG. 17. Elastic Young's modulus of Goodrich Thorodin™ AQ21 at 50 psi and at 30 °C, resulting from both horizontal and vertical shifts applied to the 10 °C and −2 °C data sets. The solid line is a curve fit.

hydrostatic pressure dependence, as was verified experimentally by measuring its elastic and loss moduli at 0.35, 1.04, and 2.07 MPa (50, 150, and 300 psi, respectively), and at three temperatures (30 °C, 10 °C, and −2 °C). At each temperature, the modulus values for all three pressures were identical (within the measurement error bounds). On the other hand, the large value of Poisson's ratio is responsible for substantial dispersive effects that need to be taken into account when converting the wave speed to Young's modulus.

1. Wave speed as a function of temperature

Figure 12 shows the magnitude of the wave speed, measured using both methods, at 30 °C, 10 °C, and −2 °C. The dots represent the values directly obtained from the measurements, without taking into account the dispersion given by Eq. (28). One can see that, even though the sample used in this study has a smaller cross section than that of the Rubatex® R451N sample, its large Poisson's ratio makes it necessary to correct for the dispersion, in order to get accurate values of the modulus. This is especially evident at high frequencies. One should also note that, at a given frequency, the effect is more pronounced at higher temperatures, where the wave speed is lower. The subsequent values of Young's modulus shown in the next section are computed from the corrected wave-speed values.

2. Young's modulus as a function of temperature

Figures 13 and 14 show the elastic and loss modulus, respectively, as a function of temperature. These data are used to perform a WLF shift as explained in Sec. V A 2, resulting in the curves depicted in Figs. 15, 16, and 17, at a reference temperature of 30 °C. The horizontal shift factors are $a_{10}=9.3$ and $a_{-2}=45.0$, and the vertical ones are $\nu_{10}=+1.8\times10^{6}$ Pa and $\nu_{10}=+2.2\times10^{6}$ Pa. Again, the loss modulus is observed to be zero in the dc limit.

VI. CONCLUDING REMARKS

This article describes a new system for measuring the complex Young's modulus of compliant polymers. The system combines a new approach to the resonant bar technique, where noncontact laser measurements are performed on samples without end mass, with a wave-speed technique that significantly extends the frequency range of the experimental investigation without requiring any sample modification. The apparatus is designed for pressure measurements ranging from 0 to 2.07 MPa (300 psi) and for temperature measurements ranging from −2 °C to 50 °C. Data obtained on Rubatex® R451N and on Goodrich Thorodin™ AQ21 demonstrated the capabilities of this new method for measuring both the elastic and the loss components of Young's modulus as a function of pressure and temperature. The Goodrich Thorodin™ AQ21 data also illustrated the fact that, for materials with a large Poisson's ratio and measured at high frequencies, dispersive effects might be present and need to be taken into account. Finally, it was shown that the temperature data collected with the system can be shifted according to the time–temperature superposition principle, to produce master curves over extended frequency ranges.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research, Code 334, Stephen Schreppler. RBX Industries Inc. is gratefully acknowledged for providing the Rubatex® R451N material.

[1] D. M. Norris, Jr. and W. C. Young, "Complex-modulus measurement by longitudinal vibration testing," Exp. Mech. **10**, 93–96 (1970).
[2] D. M. Norris, Jr. and W. C. Young, "Longitudinal forced vibration of viscoelastic bars with end mass," U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, NH 03775, Spec. Rep. 135 (1970).
[3] W. M. Madigoski and G. Lee, "Automated dynamic Young's modulus and loss factor measurements," J. Acoust. Soc. Am. **66**, 345–349 (1979).
[4] W. M. Madigoski and G. F. Lee, "Improved resonance technique for materials characterization," J. Acoust. Soc. Am. **73**, 1374–1377 (1983).
[5] M. L. Williams, R. F. Landel, and J. D. Ferry, "The temperature dependence of relaxation mechanisms in amorphous polymers and other glass-forming liquids," J. Am. Chem. Soc. **77**, 3701–3706 (1955).
[6] ANSI S2.22-1998, "Resonance method for measuring the dynamic mechanical properties of viscoelastic materials," American National Standard Institute, published through the Acoustical Society of America, New York, NY (1998).
[7] S. L. Garrett, "Resonant acoustic determination of elastic moduli," J. Acoust. Soc. Am. **88**, 210–221 (1990).
[8] Q. Guo and D. A. Brown, "Determination of the dynamic elastic moduli and internal friction using thin resonant bars," J. Acoust. Soc. Am. **108**, 167–174 (2000).
[9] R. L. Willis, L. Wu, and Y. H. Berthelot, "Determination of the complex Young and shear dynamic moduli of viscoelastic materials," J. Acoust. Soc. Am. **109**, 611–621 (2001).
[10] R. N. Capps, "Elastomeric materials for acoustical applications," Naval Research Laboratory, Underwater Sound Reference Detachment, Orlando, FL 32856 (1989).
[11] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity* (Dover, New York, 1944), p. 428.
[12] K. F. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1991), pp. 116–120.
[13] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1990), pp. 242–246.
[14] J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. 1, pp. 251–252.
[15] J. D. Ferry, *Viscoelastic Properties of Polymers*, 3rd ed. (Wiley, New York, 1980), pp. 266–270.
[16] P. H. Mott, C. M. Roland, and R. D. Corsaro, "Acoustic and dynamic mechanical properties of a polyurethane rubber," J. Acoust. Soc. Am. **111**, 1782–1790 (2002).

Behavior of rigid-porous layers at high levels of continuous acoustic excitation: Theory and experiment

O. Umnova, K. Attenborough,[a] E. Standley, and A. Cummings
Department of Engineering, University of Hull, Cottingham Road, Hull HU7 6RX, United Kingdom

(Received 5 February 2003; revised 20 June 2003; accepted 23 June 2003)

A model for the propagation of high amplitude continuous sound through hard-backed rigid-porous layers has been developed which allows for Forchheimer's correction to Darcy's law. The nonlinearity associated with this is shown to be particularly important in the range of frequencies around layer resonance. The model is based on the introduction of particle velocity dependent flow resistivity into the equivalent fluid model expression for complex tortuosity. Thermal effects are accounted for by means of a linear complex compressibility function. The model has been used to derive analytical expressions for surface impedance and reflection coefficient as a function of incident pressure amplitude. Depending on the material parameters, sample thickness, and frequency range the model predicts either growth or decrease of reflection coefficient with sound amplitude. Good agreement between model predictions and data for three rigid-porous materials is demonstrated. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1603236]

PACS numbers: 43.25.Ed, 43.25.Jh, 43.25.Ba [MFH]

LIST OF SYMBOLS

ω	Angular sound frequency
p	Pressure in sound wave
v	Particle velocity in sound wave
d	Porous sample thickness
c_0	Sound speed in air
ρ_0	Density of air
η	Viscosity of air
γ	Adiabatic constant
N_{Pr}	Prandtl number
$\delta(\omega)=\sqrt{2\eta/\omega\rho_0}$	viscous layer thickness
$\alpha(\omega)$	Complex tortuosity
$C(\omega)$	Complex compressibility
α_∞	Tortuosity
ϕ	Volume porosity
σ	dc flow resistivity
ξ	Forchheimer's nonlinearity parameter
Λ	Characteristic viscous length
k_0'	Thermal permeability
Λ'	Characteristic thermal length
$\Theta\approx 0.675(1-\phi)$	Cell model parameter
$k(\omega)=k'(\omega)+ik''(\omega)$	Wave number
$v_{\pm}(x)$	Slow varying amplitudes of particle velocity in direct and backward propagating waves inside the pores respectively
$p_{\pm}(x)$	Slow varying amplitudes of pressure in direct and backward propagating waves inside the pores, respectively
$v_{i,r}$	Particle velocity amplitudes in incident and reflected waves in air
$p_{i,r}=\pm v_{i,r}\rho_0 c_0$	Pressure amplitudes in incident and reflected waves in air
$Z_0(\omega)=Z_0'(\omega)+iZ_0''(\omega)$	Relative characteristic impedance of porous material
$Z(\omega)$	Relative surface impedance of hard backed porous sample
$R(\omega)$	Absolute value of the reflection coefficient of hard backed porous sample
ω_{res}	Resonant frequency of hard backed porous sample

I. INTRODUCTION

Acoustical properties of rigid porous media at high levels of continuous sound have been studied for more than a decade.[1–6] It has been shown that due to strong dispersion in such a medium the shock generation in an initially sinusoidal wave is mostly suppressed.[1] The dominating nonlinearity associated with growth of flow resistivity with flow velocity leads to an excess sound attenuation inside the material which has been observed in highly porous fibrous materials and porous ceramics. Forchheimer's correction has been used to extend a rigid frame theory (assuming cylindrical pores) of sound propagation in porous materials by adding a velocity dependent term to the imaginary part of complex density.[2] An interesting result was found that at some frequencies the absorption coefficient of a 9.25 cm hard-backed layer of open-cell foam first increases with pressure, peaks, and then drops off. This result was confirmed numerically. An analytical model for nonlinear surface and flow impedance, which takes account only of a nonlinear term in the imaginary part of complex density, and a numerical model, which accounts of nonlinearity in both real and imaginary parts, have been presented[4–5] and tested using a standing

[a]Author to whom correspondence should be addressed; electronic mail: k.attenborough@hull.ac.uk

wave method.[3] Good agreement between numerical solutions and surface (or flow) impedance data was demonstrated after fitting the data for the nonlinear parameters. A different approach[6] has used the equivalent fluid model of Johnson[7] with effective flow resistivity values dependent on Reynold's number to explain the nonlinear behavior of rigid porous materials. Thermal effects were included by means of the complex compressibility function introduced by Champoux and Allard[8] and Lafarge.[9] In general the resulting model requires six linear parameters (porosity, tortuosity, dc flow resistivity, characteristic viscous length, thermal permeability, and thermal length) and nonlinear parameters derived from measurement of the dc flow resistivity dependence on Reynold's number (or velocity). This approach[6] assumes that flow through the sample is constant. While this assumption was applicable to the experimental situation described in the paper, i.e., transmission measurements on relatively thin samples, it cannot be used for hard backed layers because the particle velocity in the pores is not uniform and has zero value at the impervious backing. The present paper offers a development of this approach based on the ''nonlinearization'' of the equivalent fluid model. The resulting nonlinear model enables prediction of the acoustical properties of hard backed porous layers at high intensities and includes changes of the particle velocity in the pores.

It has been established that the dependence of flow resistivity on flow velocity is quadratic when Reynolds number [defined as $\mathrm{Re}=av\rho_0/\eta$, where a is characteristic pore size, v is flow velocity in pores, ρ_0 is air density, η is air viscosity] is less than one and is linear (Forchheimer's law) for higher Reynolds numbers.[10] However in the present paper only Forchheimer's correction is considered and the initial quadratic dependence of flow resistivity on velocity is neglected. This assumption is shown to be valid for the materials used in the experiment which have relatively big (i.e., bigger than 1 mm) pores.

In Sec. II A the model for nonlinear impedance and reflection coefficient is described. It allows predictions for the whole range of frequencies. It is based on modification of an equivalent fluid model to include a flow velocity-dependent flow resistivity. As a result of this an analytical expression for the dependence of the reflection coefficient and surface impedance of a layer on the incident pressure amplitude is obtained.

In Sec. II B particular attention is given to the predicted change of reflection coefficient with incident sound pressure level near layer resonance where nonlinear effects are the strongest. An analytical criterion is derived which distinguishes between two types of behavior based on linear material parameters and layer thickness.

In Sec. III, results are compared with data. Surface admittance and reflection coefficient have been measured at different levels of continuous acoustic excitation for hard-backed layers of lead balls, porous aluminum, and porous concrete. Data are compared with the nonlinear model predictions at different frequencies, particularly those near the layer resonances.

II. THE MODEL

A. Surface impedance dependence on sound amplitude

By introducing dimensionless complex tortuosity $\alpha(\omega)$ and complex compressibility $C(\omega)$ of the fluid contained in rigid porous material, the equations for pressure p and particle velocity v variations in the plane sound wave of angular frequency ω can be written in the following form:

$$-i\omega\alpha(\omega)\rho_0 v=-\frac{dp}{dx},\quad -i\omega\frac{C(\omega)}{\rho_0 c_0^2}p=-\frac{dv}{dx}, \tag{1}$$

here c_0 is sound speed.

The dynamic (complex) tortuosity $\alpha(\omega)$ is approximated by the following function:[7]

$$\alpha(\omega)=\alpha_\infty+\frac{\sigma\phi}{-i\omega\rho_0}\sqrt{1+(-i\omega)\frac{4\alpha_\infty^2\rho_0\eta}{\sigma^2\phi^2\Lambda^2}}, \tag{2}$$

where α_∞ is the tortuosity, σ is the flow resistivity, Λ is the characteristic viscous length, ϕ is the volume porosity, and η is the coefficient of dynamic viscosity for air. The drag parameters (α_∞, σ, and Λ) for granular materials can be estimated if the particle radius and porosity are known.[11]

Similarly, the dynamic compressibility $C(\omega)$ is approximated by[8,9]

$$C(\omega)=\gamma-\frac{\gamma-1}{\left(1+\dfrac{\eta\phi}{-i\omega\rho_0 k_0' N_{\mathrm{Pr}}}\sqrt{1+(-i\omega)\dfrac{\rho_0 4k_0'^2 N_{\mathrm{Pr}}}{\eta\Lambda'^2\phi^2}}\right)}, \tag{3}$$

where γ is the adiabatic constant, N_{Pr} is the Prandtl number, k_0' is the thermal permeability, and Λ' is the characteristic thermal length. The number of parameters necessary to model sound propagation in materials with straight cylindrical pores or in granular materials can be reduced by using a similarity relationship between complex density and complex compressibility functions.[12,13] For a packing of spheres, the relationship has the following form.[13]

$$C(\omega)=\gamma-\frac{\gamma-1}{(\frac{2}{3}(1-\Theta)(\alpha(\omega N_{\mathrm{Pr}})-\alpha_\infty)+1)}, \tag{4}$$

where $\Theta\approx 0.675(1-\phi)$. It is equivalent to the following relationships between thermal permeability and dc flow resistivity:

$$k_0'=\frac{3}{2(1-\Theta)}\frac{\eta}{\sigma}, \tag{5}$$

and between viscous and thermal characteristic lengths:

$$\Lambda'=\frac{3}{2\alpha_\infty(1-\Theta)}\Lambda. \tag{6}$$

It is assumed that flow resistivity grows linearly with particle velocity amplitude in the sound wave:

$$\sigma=\sigma_0(1+\xi\phi|v|), \tag{7}$$

where ξ is Forchheimer's nonlinearity parameter which can be measured in standard flow resistivity tests. The factor of porosity in Eq. (7) allows for the difference between the velocity in pores and flow velocity measured outside the material in flow resistivity tests.

Expression (7) is substituted in Eq. (2) to get the velocity dependent complex tortuosity $\alpha(\omega,v)$:

$$\alpha(\omega,v)=\alpha_\infty+\frac{\sigma_0\phi(1+\xi\phi|v|)}{-i\omega\rho_0}\times\sqrt{1+(-i\omega)\frac{4\alpha_\infty^2\rho_0\eta}{\sigma_0^2\phi^2\Lambda^2(1+\xi\phi|v|^2)}}.$$

The solution of Eq. (1) is performed by the method of slow varying amplitudes. This method assumes that the changes in wave amplitude due to the presence of nonlinearity remain small when wave travels a distance comparable to its wavelength. Consequently the following derivations are valid only at relatively low values of incident sound pressure. First the dispersion relation is found in linear approximation

$$k(\omega)=\frac{\omega}{c_0}\sqrt{\alpha(\omega)C(\omega)}. \tag{8}$$

Then a solution of nonlinear equations is sought in the form

$$v(x)=v_+(x)e^{ik(\omega)x}+v_-(x)e^{-ik(\omega)x},$$
$$p(x)=\frac{\rho_0c_0^2}{i\omega C(\omega)}\frac{dv}{dx}, \tag{9}$$

where $v_\pm(x)$ are the slow varying amplitudes of particle velocity in the forward and backward propagating waves. Expression (9) is now substituted into Eq. (1) with second derivatives of $v_\pm(x)$ neglected. This leads to a pair of first-order nonlinear differential equations describing spatial changes in $v_\pm$:

$$\begin{aligned}\frac{dv_+}{dx}&=-M(\omega)\left(\sqrt{1+i\frac{\omega_c}{\omega}(1+\xi\phi|v|)^2}-\sqrt{1+i\frac{\omega_c}{\omega}}\right)v_+,\\ \frac{dv_-}{dx}&=M(\omega)\left(\sqrt{1+i\frac{\omega_c}{\omega}(1+\xi\phi|v|)^2}-\sqrt{1+i\frac{\omega_c}{\omega}}\right)v_-,\end{aligned} \tag{10}$$

where

$$M(\omega)=(1-i)\frac{\alpha_\infty\omega^2C(\omega)}{c_0^2k(\omega)}\frac{\delta(\omega)}{2\Lambda},\quad \omega_c=\frac{\sigma_0^2\phi^2\Lambda^2}{4\alpha_\infty^2\rho_0\eta},$$

and $\delta(\omega)=\sqrt{2\eta/\omega\rho_0}$ is the viscous layer thickness.

To determine the normalized surface impedance of the layer $Z=(1/\rho_0c_0)[p(0)/\phi v(0)]$, Eq. (10) has to be solved with boundary condition of flux continuity and pressure continuity on the material surface

$$v_i+v_r=\frac{p_i-p_r}{\rho_0c_0}=\phi v(0)=\phi(v_+(0)+v_-(0)), \tag{11}$$

$$p_i+p_r=p(0)=\rho_0c_0\sqrt{\frac{\alpha(\omega)}{C(\omega)}}\left(v_+(0)-v_-(0)+\frac{1}{ik(\omega)}\left(\frac{dv_+}{dx}+\frac{dv_-}{dx}\right)\Big|_{x=0}\right) \tag{12}$$

and the condition of zero particle velocity on an impervious hard backing:

$$v(d)=v_+(d)e^{ik(\omega)d}+v(d)e^{-ik(\omega)d}=0, \tag{13}$$

where d is porous layer thickness. The amplitude of the reflection coefficient follows from

$$R=\left|\frac{p_r}{p_i}\right|=\left|\frac{1-Z}{1+Z}\right|,$$

where p_i and p_r are pressure amplitudes in the incident and reflected waves, respectively.

This is equivalent to the following expression (which will be used later):

$$R=\sqrt{1-4\frac{\mathrm{Re}(Z)}{1+ZZ^*+2\,\mathrm{Re}(Z)}}. \tag{14}$$

The solution of Eq. (10) has been performed using the mean-field approximation.[14] This leads to the following relationships between the velocity amplitudes on both sides of the layer:

$$v_+(d)\approx v_+(0)\exp\left[-M(\omega)\left(\frac{\sqrt{1+i\frac{\omega_c}{\omega}(1+\xi\phi|v(0)|)^2}+\sqrt{1+i\frac{\omega_c}{\omega}(1+\xi\phi|v(d)|)^2}}{2}-\sqrt{1+i\frac{\omega_c}{\omega}}\right)d\right], \tag{15}$$

$$v_-(d)=v_-(0)\frac{v_+(0)}{v_+(d)}.$$

By combining this with boundary conditions (11)–(13) the following transcendental equation for the dependence of surface impedance on incident pressure amplitude has been obtained:

$$Z(\omega)=\frac{1}{\phi}\sqrt{\frac{\alpha(\omega)}{C(\omega)}}\left(1-\frac{M(\omega)}{ik(\omega)}G(U(\omega))\right)\times\coth\left(-ik(\omega)d\left(1-\frac{M(\omega)}{2ik(\omega)}G(U(\omega))\right)\right),$$
$$U(\omega)=\frac{2|p_i|}{\phi|1+Z(\omega)|\rho_0 c_0}, \tag{16}$$

where

$$G(x)=\sqrt{1+i\frac{\omega_c}{\omega}(1+\xi\phi x)^2}-\sqrt{1+i\frac{\omega_c}{\omega}}.$$

B. Reflection coefficient at resonance: Dependence on sound amplitude

The following derivations are valid for

$$\frac{\omega_c}{\omega}<1, \tag{17}$$

i.e., for frequencies higher than ω_c. It will be shown in Sec. III that the approximation is valid for the layer resonance frequencies of the tested samples.

The following expressions are used to approximate the complex density and complex compressibility functions in this frequency range:

$$\alpha(\omega)\approx\alpha_\infty\left(1+\frac{(1+i)\delta(\omega)}{\Lambda}\left(1+\frac{\omega_c}{-2i\omega}\right)\right),$$
$$C(\omega)\approx 1+\frac{\gamma-1}{\sqrt{N_{\mathrm{Pr}}}}\frac{(1+i)\delta(\omega)}{\Lambda'}.$$

Consequently the propagation constant k and characteristic impedance Z_0 can be approximated by

$$k(\omega)\approx\frac{\sqrt{\alpha_\infty}\omega}{c_0}\left[\left(1+\frac{\delta(\omega)}{2L_1}-\frac{\delta(\omega)}{2\Lambda}\frac{\omega_c}{2\omega}\right)+i\left(\frac{\delta(\omega)}{2L_1}+\frac{\delta(\omega)}{2\Lambda}\frac{\omega_c}{2\omega}\right)\right],$$
$$Z_0(\omega)\approx\frac{\sqrt{\alpha_\infty}}{\phi}\left[\left(1+\frac{\delta(\omega)}{2L_2}-\frac{\delta(\omega)}{2\Lambda}\frac{\omega_c}{2\omega}\right)+i\left(\frac{\delta(\omega)}{2L_2}+\frac{\delta(\omega)}{2\Lambda}\frac{\omega_c}{2\omega}\right)\right],$$

where $L_{1,2}$ are functions of the characteristic lengths given by

$$\frac{1}{L_{1,2}}=\frac{1}{\Lambda}\pm\frac{\gamma-1}{\sqrt{N_{\mathrm{Pr}}}\Lambda'}.$$

It is assumed in the following derivation that in the frequency range of interest the viscous layer depth is small compared with L_2,

$$\frac{\sqrt{\alpha_\infty}}{2\phi}\frac{\delta(\omega)}{L_2}<1. \tag{18}$$

By using expression (14) the following approximate expression for the reflection coefficient (linear) can be derived:

$$R(\omega)=\sqrt{1-4\frac{Z_0'(\omega)(1-e^{-4k''(\omega)d})}{(1+Z_0'(\omega)^2)(1+e^{-4k''(\omega)d})+2Z_0'(\omega)(1-e^{-4k''(\omega)d})+2e^{-4k''(\omega)d}(Z_0'(\omega)^2-1)\cos(2k'(\omega)d)}},$$

where $k'(\omega)=\mathrm{Re}(k(\omega))$, $k''(\omega)=\mathrm{Im}(k(\omega))$, $Z_0'(\omega)=\mathrm{Re}(Z_0(\omega))$.

In the approximation adopted here the real part of propagation constant $k'(\omega)$ shows the strongest frequency dependence of all functions involved in $R(\omega)$. Due to this fact, the first resonant frequency ω_{res} where reflection coefficient reaches its minimum can be found approximately from the condition:

$$2k'(\omega_{\mathrm{res}})d=\pi.$$

Retaining only the linear in ω_c/ω terms the following expression for the resonant frequency is derived:

$$\omega_{\mathrm{res}}=\omega_{\mathrm{res}}^{(0)}\left[1+\frac{\omega_c}{2\omega_{\mathrm{res}}^{(0)}}\frac{\delta_{\mathrm{res}}^{(0)}}{2\Lambda}\frac{1}{1+\frac{\delta_{\mathrm{res}}^{(0)}}{4L_1}}\right], \tag{19}$$

where

$$\delta_{\mathrm{res}}^{(0)}=\delta(\omega_{\mathrm{res}}^{(0)}),$$
$$\omega_{\mathrm{res}}^{(0)}=\frac{1}{4}\left(\sqrt{\frac{2\pi c}{\sqrt{\alpha_\infty}d}+\frac{\eta}{2\rho_0 L_1^2}}-\sqrt{\frac{\eta}{2\rho_0 L_1^2}}\right)^2.$$

The value of the reflection coefficient at this frequency may be approximated by

$$R_{\mathrm{res}}=R(\omega_{\mathrm{res}})\approx\sqrt{1-4\frac{\sqrt{\alpha_\infty}}{\phi}\frac{\left(1-e^{-2n}\exp\left(-2l\left(\frac{\omega_c}{\omega_{\mathrm{res}}^{(0)}}\right)\right)\right)}{\left(\left(1+\frac{\sqrt{\alpha_\infty}}{\phi}\right)-e^{-n}\exp\left(-l\left(\frac{\omega_c}{\omega_{\mathrm{res}}^{(0)}}\right)\right)\right)\left(\frac{\sqrt{\alpha_\infty}}{\phi}-1\right)\right)^2}},\tag{20}$$

where

$$n=\frac{\delta_{\mathrm{res}}^{(0)}}{L_1}\frac{\sqrt{\alpha_\infty}\omega_{\mathrm{res}}^{(0)}}{c_0}d,\quad l=n\frac{L_1}{\Lambda}.$$

The above-presented results have been obtained assuming that sound amplitude is low and material behaves linearly. To obtain the approximate dependence of the resonant value of reflection coefficient on incident sound amplitude we have to replace the parameter ω_c in Eq. (20) with an effective value that depends on the sound amplitude. Taking into account the results of the previous section, at resonance ω_c can be approximated by

$$\omega_c^{\mathrm{eff}}=\omega_c\left(1+\xi\frac{2|p_i|}{\rho_0c_0|1+Z_{\mathrm{res}}|}\right)^2,$$

where Z_{res} is the (linear) impedance at the layer resonance frequency.

The first derivative of R_{res} with respect to p_i is

$$\frac{dR_{\mathrm{res}}(|p_i|)}{d(|p_i|)}=-A\left[\left(1+\frac{\sqrt{\alpha_\infty}}{\phi}\right)e^{-n}\exp\left(-l\left(\frac{\omega_c^{\mathrm{eff}}}{\omega_{\mathrm{res}}^{(0)}}\right)\right)-\left(\frac{\sqrt{\alpha_\infty}}{\phi}-1\right)\right],$$

where A is a positive constant.

This means that reflection coefficient at resonance always grows with incident pressure amplitude as long as the following inequality is valid:

$$\left(1+\xi\frac{2|p_i|}{\rho_0c_0|1+Z_{\mathrm{res}}|}\right)^2>\frac{\omega_{\mathrm{res}}^{(0)}}{\omega_c l}\left(\ln\left(\frac{1+\frac{\sqrt{\alpha_\infty}}{\phi}}{\frac{\sqrt{\alpha_\infty}}{\phi}-1}\right)-n\right).\tag{21}$$

This suggests that two types of behavior are possible depending on the material parameters. If the right-hand side of inequality (21) is less than one, i.e., if

$$\omega_c>\frac{\Lambda c_0}{\sqrt{\alpha_\infty}\delta_{\mathrm{res}}^{(0)}d}\ln\left(\frac{1+\frac{\sqrt{\alpha_\infty}}{\phi}}{\sqrt{\frac{\alpha_\infty}{\phi}-1}}\right)-\omega_{\mathrm{res}}^{(0)}\frac{\Lambda}{L_1},\tag{22}$$

then the reflection coefficient at layer resonance grows as the incident pressure increases. [Fig. 1(a)]. This behavior is characteristic for fairly resistive materials. For materials with parameters such that Eq. (22) is not obeyed, the reflection coefficient decreases when pressure amplitude is less than some critical value p_{cr}. For higher amplitudes, the initial decrease is followed by subsequent growth [Fig. 1(b)]. This behavior is predicted for low flow resistivity materials.

It follows from Eq. (21) that p_{cr} is higher for materials with weak nonlinearity (i.e., relatively low values of parameter ξ). Note that due to the approximate nature of complex tortuosity function (2) only the two first terms in its high frequency expansion are exact. Consequently, criterion (22) should not be treated as a precise result.

III. COMPARISON WITH DATA

A. Lead shot

Impedance measurements have been carried out, in 5.1 cm (inner) diameter tubes connected to a 400 W Fane loudspeaker using the transfer function method.[3] A vertical 70-cm-long tube was used for measurements on granular materials, and a horizontal 3-m-long tube was used for rigid porous materials. Two $\frac{1}{4}$ in. Bruel & Kjaer microphones were placed 20 cm apart in the tube. The distance between the sample surface and the closest microphone was about 10 cm. A metallic sample holder with an acoustically transparent mesh designed to fit the vertical impedance tube was used

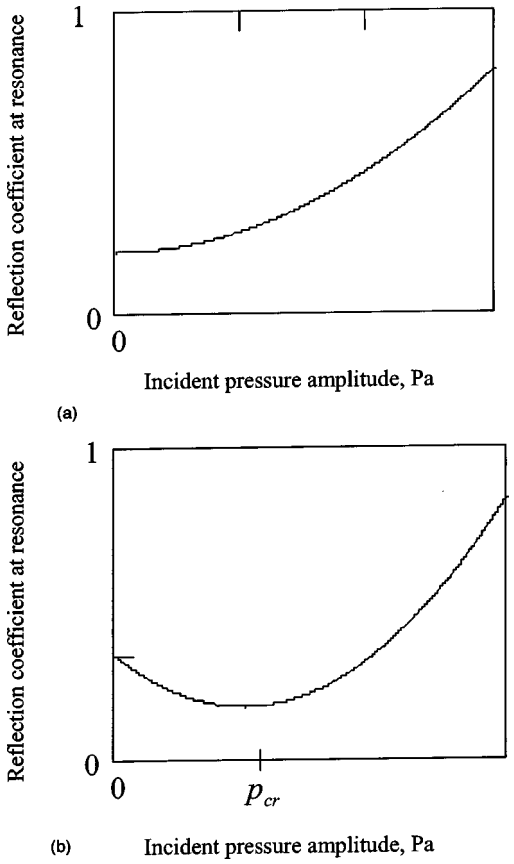


FIG. 1. The two types of reflection coefficient dependence on the incident pressure amplitude predicted by the model.

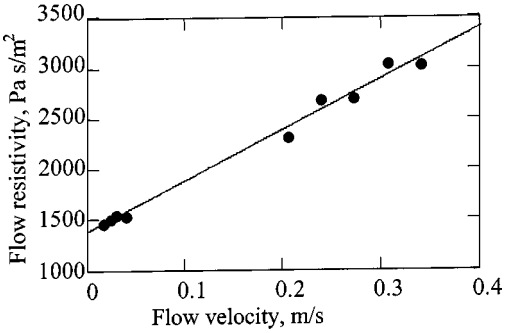


FIG. 2. The dependence of flow resistivity on flow velocity measured for lead shot and its straight line approximation.

for the unconsolidated materials. The porous aluminum and porous concrete samples were specially manufactured to fit in the horizontally installed tube. To investigate possible interference of nonlinear effects inside the tube, the transfer function has been measured in the absence of samples and with a rigid termination. It has proved to be independent of amplitude.

The flow resistivities of the tested rigid-porous materials have been deduced from measurements of volume velocity and pressure drop across samples of known thickness subject to compressed air in a tube. As with the impedance measurements, a vertical sample holder was used for measurements on loose materials and a horizontal one for rigid-porous materials. The inner diameter of the flow resistivity tubes was close to 50 mm.

The model has been tested against impedance data for a 15 cm layer of lead shot with mean radius of 1.89 mm. This material consists of nearly identical spherical particles and shows strong $\sigma(v)$ dependence even at relatively low values of flow velocity (Fig. 2) but the dependence is approximately linear. This approximation was used to determine Forchheimer's nonlinearity parameter ξ and dc flow resistivity value σ_0. The value of dc flow resistivity (1373 Pa s/m^2) deduced from data is fairly close to that predicted (1465 Pa s/m^2) by a cell model for identical spherical particles.[11] The characteristic viscous length was estimated by using the cell model also. The tortuosity was estimated using the empirical formula $\alpha=1/\sqrt{\phi}\approx 1.6$[15] as use of the value (1.8) predicted by the cell model[11] gives slightly incorrect values for the layer resonance frequencies. Values of parameters of the lead shot packing are given in Table I. Relationship (4) has been used to determine the complex compressibility function. Predictions of the linear equivalent fluid model are compared with

TABLE I. Parameters for lead shot and their method of determination.

Porosity ϕ	0.385	Measured
Flow resistivity σ_0 (Pa s/m^2)	1373	Measured
Characteristic viscous length Λ (m)	5.5×10^{-4}	Cell model[a]
Tortuosity α_∞	1.6	Empirical formula[b]
Forchheimer's parameter ξ (s/m)	3.7	Measured

[a]Reference 11.
[b]Reference 15.

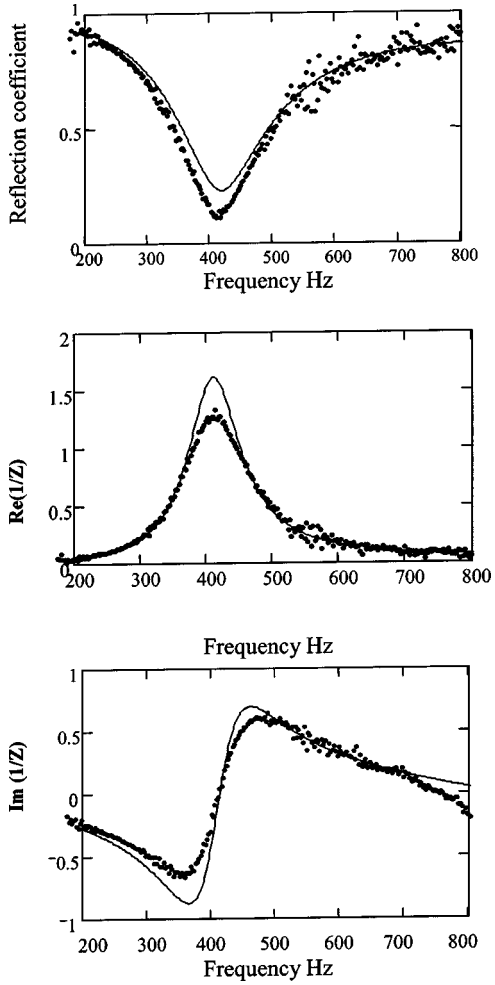


FIG. 3. The frequency dependence of reflection coefficient and surface admittance for low amplitude sound for lead shot. Points—white noise data, line—Johnson/Champoux/Allard/Lafarge linear model.

data in Fig. 3. It should be noted that the cell model has a tendency to overestimate the characteristic viscous length value for denser packings.[11] This might explain the slight disagreement between theory and data. The reflection coefficient of a layer achieves its minimum value of 0.11 at a frequency of approximately 415 Hz. In the resonant region, both inequalities (17) and (18) are valid: $\omega_c/\omega_{\rm res}\approx 0.2$, $(\sqrt{\alpha_\infty}/2\phi)[\delta(\omega_{\rm res})/L_2]\approx 0.2$. This means that the results of previous section can be used to describe any nonlinear behavior at resonance. Indeed the resonant frequency predicted by Eq. (19) is 403 Hz, which is close to that shown in the data. The value of the reflection coefficient at resonance predicted by Eq. (20) is approximately 0.2, which is very close to the value (0.23) given by the linear equivalent fluid model. However both are higher than the measured value for the above-mentioned reasons.

Since, for this material, condition (22) is not satisfied, i.e.,

$$\omega_c\approx 378\ \text{Hz},$$

$$\frac{\Lambda c_0}{\sqrt{\alpha_\infty}\delta^{(0)}_{\rm res}d}\ln\left(\frac{1+\dfrac{\sqrt{\alpha_\infty}}{\phi}}{\dfrac{\sqrt{\alpha_\infty}}{\phi}-1}\right)-\omega^{(0)}_{\rm res}\frac{\Lambda}{L_1}\approx 2408\ \text{Hz},$$

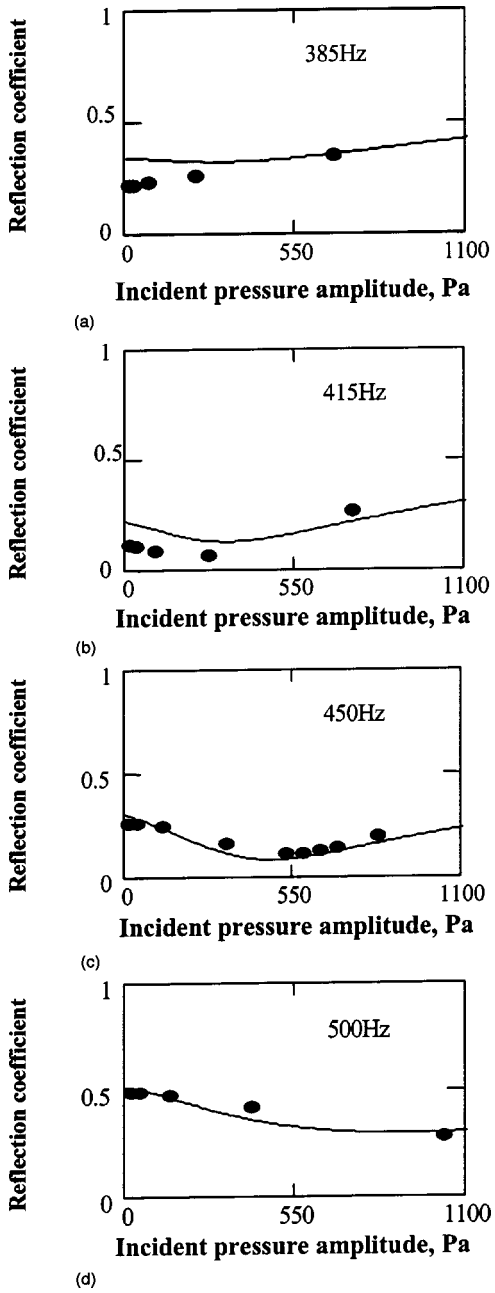


FIG. 4. The reflection coefficient as a function of the pressure amplitude of the incident wave at four different frequencies for lead shot. Points—data, lines—model. (a) 385 Hz, (b) 415 Hz, (c) 450 Hz, (d) 500 Hz.

the reflection coefficient at resonance should experience decrease followed by subsequent growth as the incident sound amplitude is increased.

To study the nonlinear behavior of the material, measurements have been carried out at different values of incident pressure amplitude using pure tone signals in the frequency range between 375 and 600 Hz.

Figures 4 and 5 show that the nonlinear model predicts the observed variation of the reflection coefficient and admittance with the incident pressure. Growth or decrease of the reflection coefficient is predicted depending on the frequency. At the first layer resonance frequency, the reflection coefficient is observed to decrease as the pressure amplitude increases to 266 Pa. As predicted by the theory, this decrease

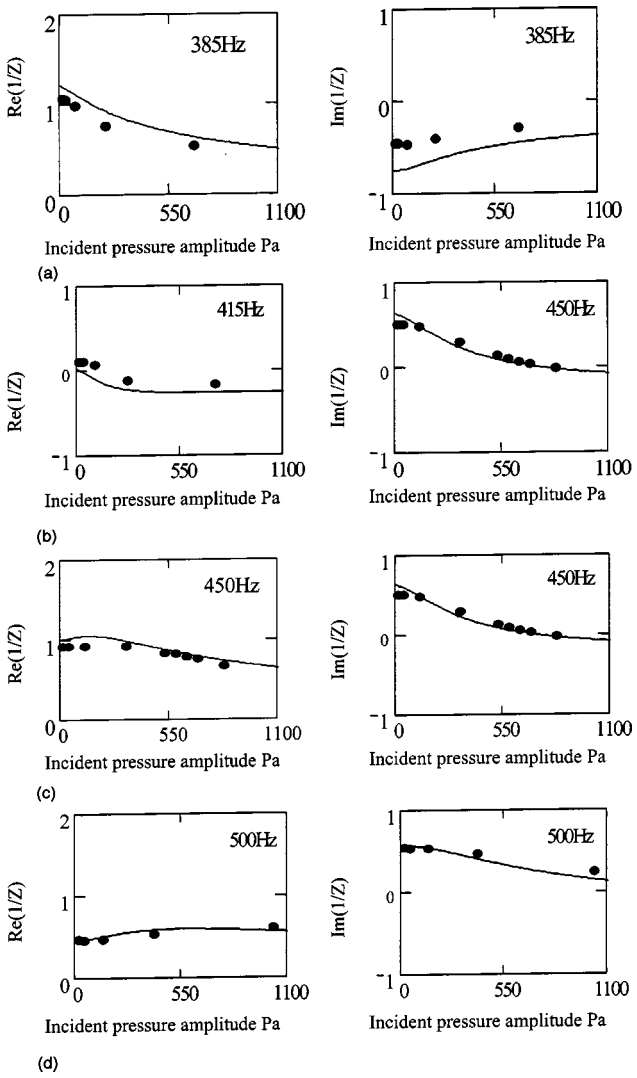


FIG. 5. The surface admittance $1/Z$ as a function of the pressure amplitude of the incident wave at four different frequencies for lead shot. Points—data, lines—model. (a) 385 Hz, (b) 415 Hz, (c) 450 Hz, (d) 500 Hz.

is followed by growth as the pressure amplitude is increased further. It should be noted that these predictions have been obtained with no adjustable parameters.

B. Porous aluminum

Impedance measurements have been made on a 15 cm sample of aluminum foam which is a highly porous rigid

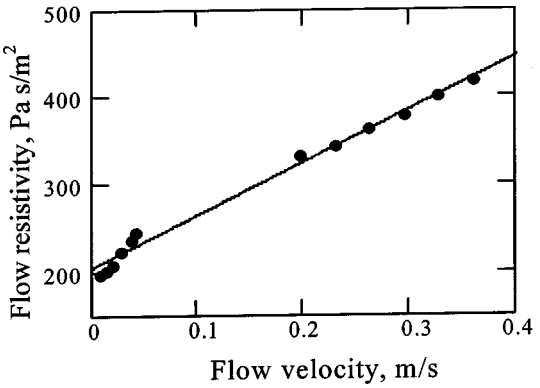


FIG. 6. The dependence of flow resistivity on flow velocity measured for porous aluminum and its straight-line approximation.

TABLE II. Parameters for porous aluminum and their method of determination.

Porosity ϕ	0.93	Measured
dc flow resistivity σ_0 (Pa s/m^2)	204.6	Measured
Characteristic viscous length Λ (m)	7.7×10^{-4}	Deduced by fitting (linear) characteristic impedance data
Tortuosity α_∞	1.07	Deduced by fitting (linear) characteristic impedance data
Forchheimer's parameter ξ (s/m)	2.9	Measured

material. The value of porosity obtained by weighing the sample is 0.93. Its flow resistivity is low and shows quite strong dependence on flow velocity (Fig. 6). The characteristic viscous length and tortuosity of the material have been deduced to fit the low intensity characteristic impedance data. To account for thermal effects it has been assumed that the rigid inclusions are spherical as in the case of lead shot. Of course this ideal microstructure does not reflect the complicated pore geometry of the foam. However, the high porosity of this material means that the thermal effects are

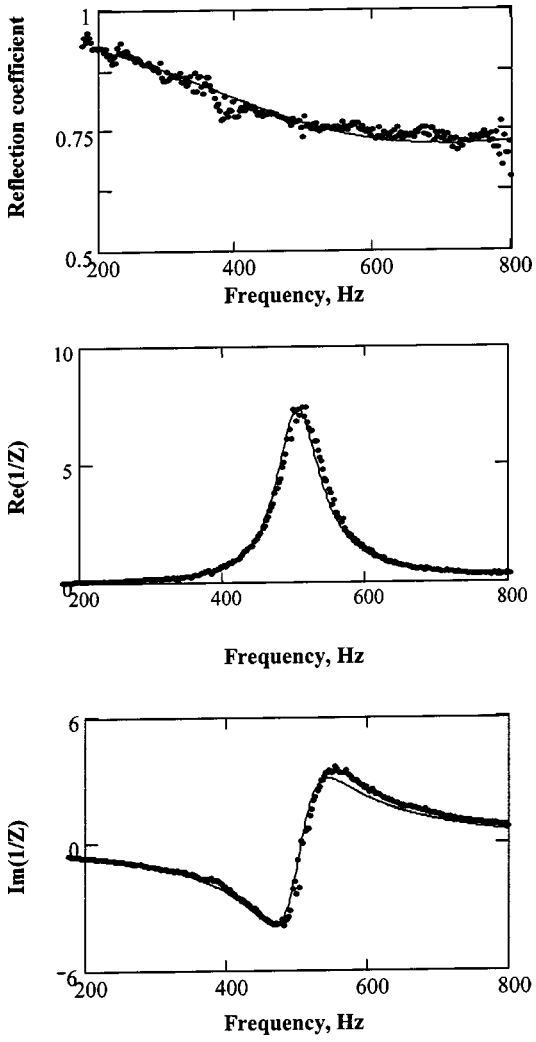


FIG. 7. The frequency dependence of the reflection coefficient and the surface admittance for low amplitude sound for porous aluminum. Points—white noise data, lines—Johnson/Champoux/Allard/Lafarge linear model.

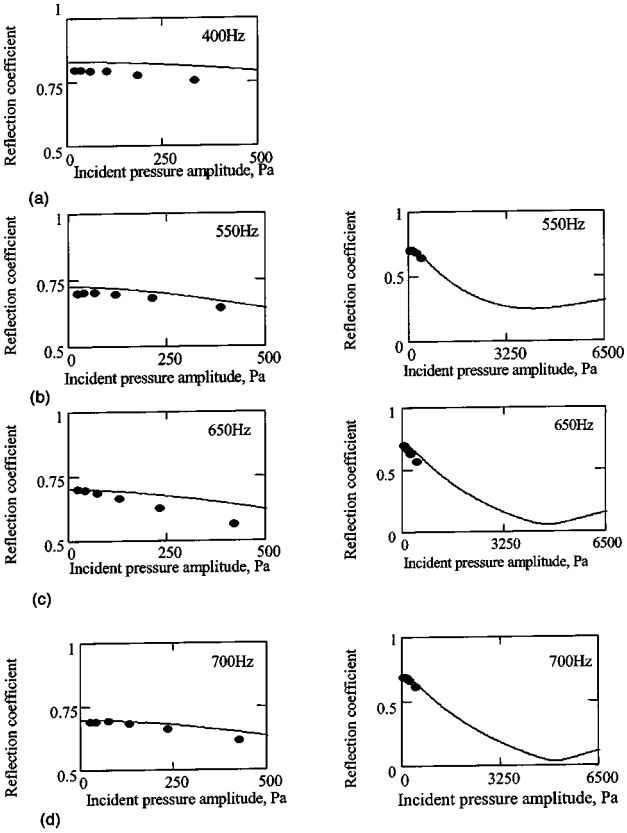


FIG. 8. The reflection coefficient of porous aluminum as a function of pressure amplitude in the incident wave at four different frequencies. Points—data, lines—model. (a) 400 Hz, (b) 550 Hz, (c) 650 Hz, (d) 700 Hz.

weak, so predictions are not much affected by use of different idealizations of the microstructure, for instance straight cylindrical pores. Values of parameters used in the modeling are summarized in Table II.

Figure 7 shows that the first layer resonance in the reflection coefficient is not very sharp and lies approximately between 600 and 700 Hz. The value of reflection coefficient at resonance is close to 0.75.

Conditions (17) and (18) are valid for the material since $\omega_c/\omega_{\rm res}\approx0.07$ and $(\sqrt{\alpha_\infty}/2\phi)[\delta(\omega_{\rm res})/L_2]\approx0.04$. The approximate resonant frequency predicted by Eq. (19) is 511 Hz, which is close to the frequency of the zero axis crossing of the measured reactance. The value of the reflection coefficient at resonance predicted by Eq. (20) is 0.75, which is close to the data. Condition (20) is not valid for porous aluminum since

$$\omega_c\approx215\ \text{Hz},$$

$$\frac{\Lambda c_0}{\sqrt{\alpha_\infty}\delta^{(0)}_{\rm res}d}\ln\left(\frac{1+\dfrac{\sqrt{\alpha_\infty}}{\phi}}{\dfrac{\sqrt{\alpha_\infty}}{\phi}-1}\right)-\omega^{(0)}_{\rm res}\frac{\Lambda}{L_1}\approx47\,060\ \text{Hz},$$

This means that growth of the incident pressure amplitude should result in reflection coefficient behavior similar to that observed in lead shot. Figures 8 and 9 show that there is a good agreement between data and model predictions for the reflection coefficient and admittance behavior. However

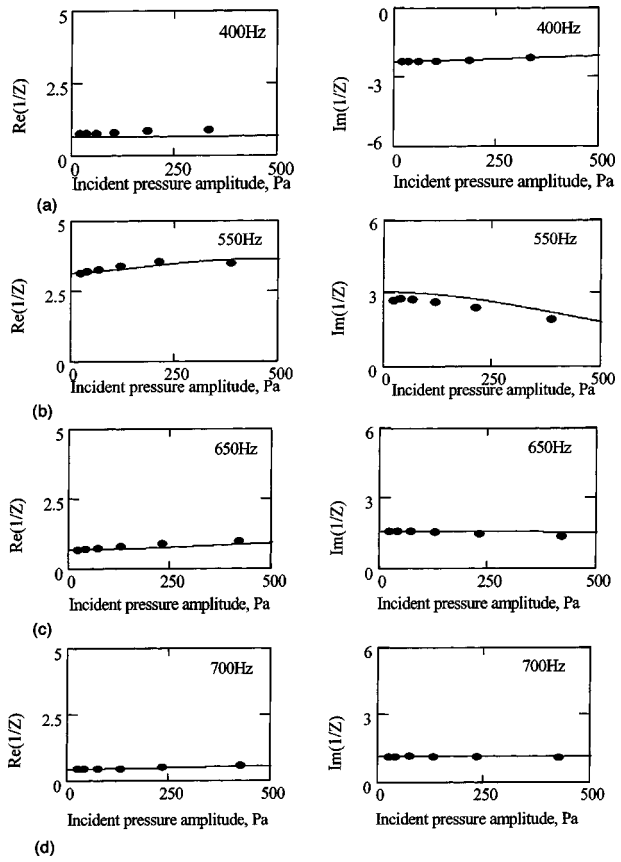


FIG. 9. The surface admittance 1/Z of porous aluminum as a function of the pressure amplitude in the incident wave at four different frequencies. Points—data, lines—model. (a) 400 Hz, (b) 550 Hz, (c) 650 Hz, (d) 700 Hz.

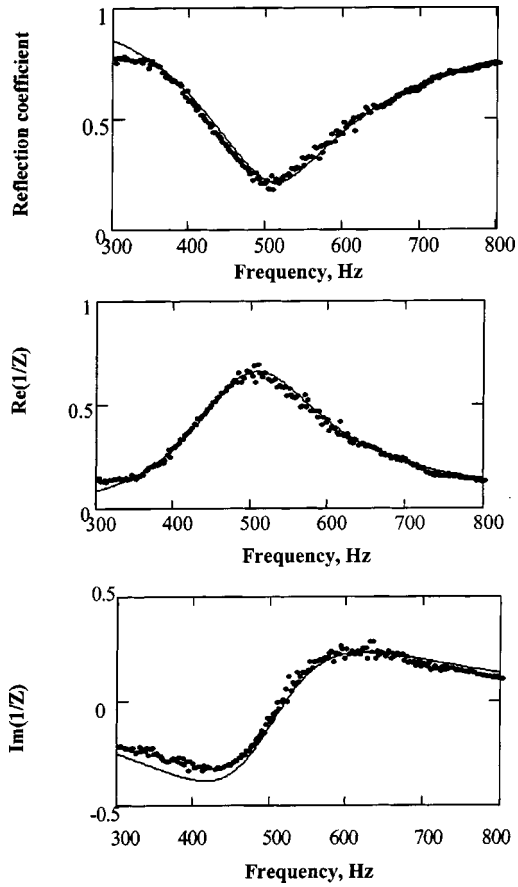


FIG. 11. The frequency dependence of the reflection coefficient and the surface admittance for low amplitude sound incident on porous concrete. Points—white noise data, lines—Johnson/Champoux/Allard/Lafarge linear model.

the levels of acoustic excitation that could be achieved in the experiments were too low to show that the decrease of the reflection coefficient with pressure amplitude is followed by its growth. By extrapolating the predictions to pressures up to 6.5 kPa it is possible to show that this should happen at amplitudes in the range of 4–5 kPa.

C. Porous concrete

A 10-cm-long cylindrical sample of porous concrete has been used as the final example for the measurements reported here. It consists of grains with irregular shape and dimensions mainly in the range of 1–6 mm held together by the binder. The flow resistivity of this material is much higher than that of lead shot and porous aluminum and the measured Forchheimer's nonlinearity is the strongest of all the materials (Fig. 10). The material porosity has been measured using a new method[16] and is close to 0.3. Characteristic viscous length and tortuosity have been fitted to get the best agreement for low intensity impedance and reflection coefficient data and equivalent fluid model predictions (Fig. 11). Thermal effects have been included by assuming identical spherical particles and using relationship (4). The parameter values are summarized in Table III. The first layer resonance value in the reflection coefficient is at approximately 500 Hz and is equal to 0.27. Both conditions (17) and (18) are valid near the resonance since $\omega_c/\omega_{\mathrm{res}} \approx 0.07$ and $(\sqrt{\alpha_\infty}/2\phi)[\delta(\omega_{\mathrm{res}})/L_2] \approx 0.7$.

The frequency of the first layer resonance predicted by

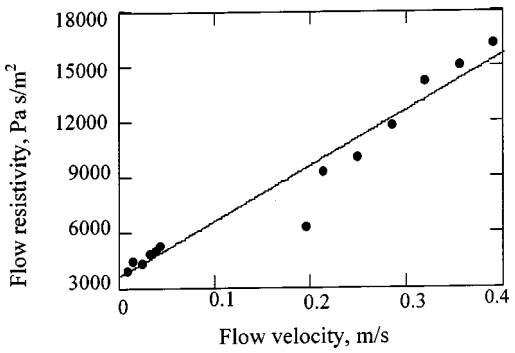


FIG. 10. The dependence of flow resistivity on flow velocity measured for porous concrete and its straight-line approximation.

TABLE III. Parameters for porous concrete and their method of determination.

Porosity ϕ	0.3	Measured
dc flow resistivity σ_0 (Pa s/m^2)	3619	Measured
Characteristic viscous length Λ (m)	2.2×10^{-4}	Deduced by fitting (linear) surface admittance data
Tortuosity α_∞	1.8	Deduced by fitting (linear) surface admittance data
Forchheimer's parameter ξ (s/m)	8.3	Measured

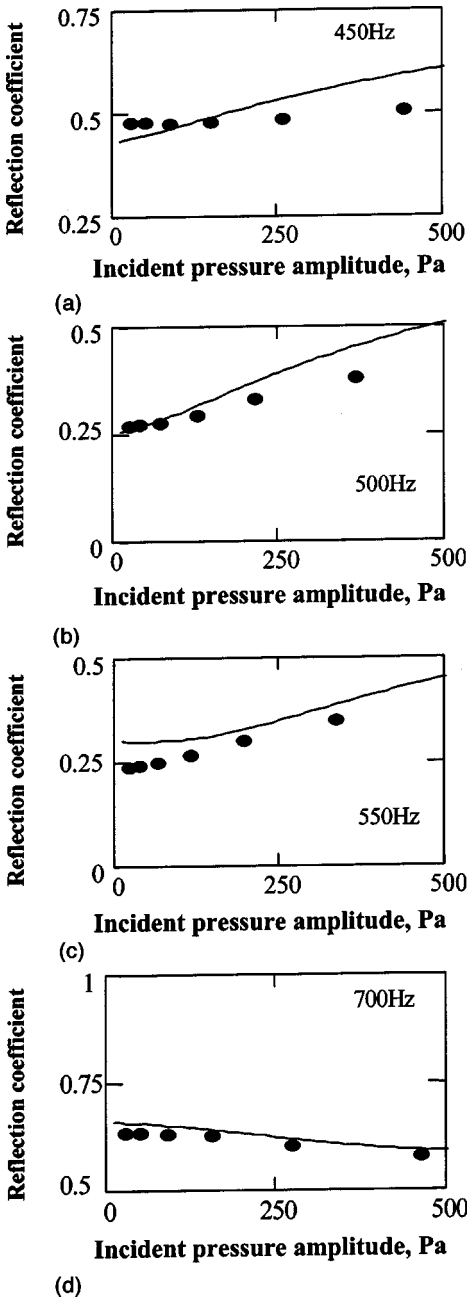


FIG. 12. The reflection coefficient of porous concrete as a function of pressure amplitude in incident wave at four different frequencies. Points—data, lines—model. (a) 450 Hz, (b) 500 Hz, (c) 550 Hz, (d) 700 Hz.

Eq. (19) is 515 Hz, which is close to that shown by the data. The value of the reflection coefficient at layer resonance predicted by Eq. (20) is 0.21, which is close to that measured also. Since, for this material, condition (21) is satisfied, i.e.,

$$\omega_c \approx 226 \text{ Hz},$$

$$\frac{\Lambda c_0}{\sqrt{\alpha_\infty}\delta_{\text{res}}^{(0)} d}\ln\left(\frac{1+\frac{\sqrt{\alpha_\infty}}{\phi}}{\frac{\sqrt{\alpha_\infty}}{\phi}-1}\right) - \omega_{\text{res}}^{(0)}\frac{\Lambda}{L_1} \approx -1365 \text{ Hz},$$

the reflection coefficient at the first layer resonance is predicted to grow with the incident pressure amplitude. This prediction is confirmed experimentally (Fig. 12). When the pressure amplitude is increased to 365 Pa, the reflection coefficient at resonance experiences growth of 40% compared with its low intensity value. Moreover there is a good quantitative agreement between theory and data for the reflection coefficient and admittance over the whole range of frequencies (Fig. 13).

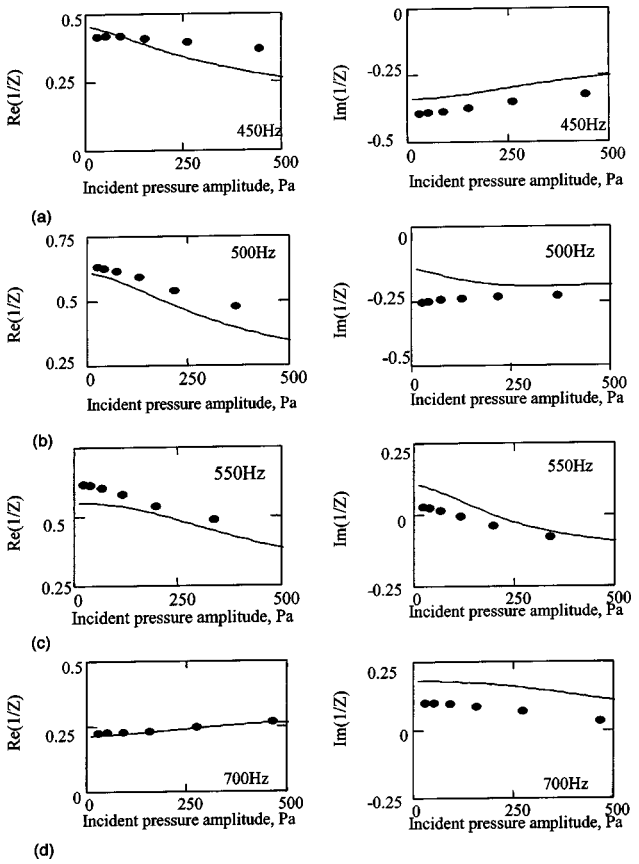


FIG. 13. Surface admittance $1/Z$ as a function of pressure amplitude in incident wave at different frequencies, porous concrete. Points—data, lines—model. (a) 450 Hz, (b) 500 Hz, (c) 550 Hz, (d) 700 Hz.

IV. CONCLUSION

A model which combines Forchheimer's nonlinearity with an equivalent fluid model has been developed and tested successfully against data for three types of hard-backed rigid-porous layers (lead shot, porous aluminum, and porous concrete). The model is shown to give valid predictions of both impedance and reflection coefficient behavior at high sound amplitudes (up to 500 Pa) over a wide range of frequencies. Particular attention has been paid to the reflection coefficient behavior near layer resonance. Depending on material parameters and layer thickness, either growth or decrease of reflection coefficient with pressure amplitude is predicted and measured. The approximate criterion derived in the paper allows predictions of the nonlinear behavior at resonance based on the material's linear parameters and on the assumption that first layer resonant frequency is relatively high. It has been tested against data for low and medium flow resistivity materials and has proved satisfactory.

ACKNOWLEDGMENTS

The work is supported in part by USARSDG (UK), Contract No. R7D 8901-EN-01 with funds from the US Army ERDC BT-25 Program. We are grateful to Marshalls plc. for supplying porous concrete samples.

[1] H. L. Kuntz and D. T. Blackstock, ''Attenuation of intense sinusoidal waves in air-saturated, bulk porous materials,'' J. Acoust. Soc. Am. **81**, 1723–1731 (1987).
[2] D. K. Wilson, J. D. McIntosh, and R. F. Lambert, ''Forchheimer-type nonlinearity for high intensity propagation of pure tones in air—saturated porous media,'' J. Acoust. Soc. Am. **84**, 350–359 (1988).
[3] J. D. McIntosh, M. T. Zorusky, and R. F. Lambert, ''Standing wave apparatus for measuring fundamental properties of acoustic materials in air,'' J. Acoust. Soc. Am. **88**, 1929–1938 (1990).
[4] J. D. McIntosh and R. F. Lambert, ''Nonlinear wave propagation through rigid porous materials. I. Nonlinear parametrization and numerical solutions,'' J. Acoust. Soc. Am. **88**, 1939–1949 (1990).
[5] R. F. Lambert and J. D. McIntosh, ''Nonlinear wave propagation through rigid porous materials. II. Approximate analytical solutions,'' J. Acoust. Soc. Am. **88**, 1950–1959 (1990).
[6] Y. Auregan and M. Pachebat, ''Measurement of the nonlinear behaviour of acoustical rigid porous materials,'' Phys. Fluids **11**, 1342–1345 (1999).
[7] D. L. Johnson, J. Koplik, and R. Dashen, ''Theory of dynamic permeability and tortuosity in fluid-saturated porous media,'' J. Fluid Mech. **176**, 379–402 (1987).
[8] Y. Champoux and J.-F. Allard, ''Dynamic tortuosity and bulk modulus in air-saturated porous media,'' J. Appl. Phys. **70**, 1975–1979 (1991).
[9] D. Lafarge, P. Lemariner, J.-F. Allard, and V. Tarnow, ''Dynamic compressibility of air in porous structures at audible frequencies,'' J. Acoust. Soc. Am. **102**, 1995–2006 (1997).
[10] M. Firdaouss, J.-L. Guermond, and P. LeQuere, ''Nonlinear correction to Darcy's law at low Reynold's numbers,'' J. Fluid Mech. **343**, 331–350 (1997).
[11] O. Umnova, K. Attenborough, and K. M. Li, ''Cell model calculations of the dynamic drag parameters in packings of spheres,'' J. Acoust. Soc. Am. **107**, 3113–3119 (2000).
[12] M. R. Stinson, ''The propagation of plane sound waves in narrow and wide circular tubes, and generalization to uniform tubes of arbitrary cross-sectional shape,'' J. Acoust. Soc. Am. **89**, 550–558 (1989).
[13] O. Umnova, K. Attenborough, and K. M. Li, ''A cell model for the acoustical properties of packings of spheres,'' Acta Acust. **87**, 226–235 (2001).
[14] H. M. Gibbs, *Optical Bistability, Controlling Light by Light* (Academic, Orlando, 1985).
[15] K. Attenborough, ''Models for the acoustical characteristics of air filled granular materials,'' Acta Acust. **1**, 213–226 (1993).
[16] P. Leclaire, O. Umnova, K. V. Horoshenkov, and L. Maillet, ''Porosity measurements by comparison of air volumes,'' Rev. Sci. Instrum. **74**, 1366–1370 (2003).

Particle drift in a resonance tube—a numerical study

A. Alexeev and C. Gutfinger[a)]
Laboratory of Aerosol Research, Faculty of Mechanical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel

(Received 14 March 2003; accepted for publication 16 June 2003)

The paper considers longitudinal drift of small particles in a resonance tube, caused by periodic shock waves, and its effect on particle agglomeration. It is found that depending on particle size, drift is caused by shock waves and/or gas acceleration and compression. It is also shown that the drift velocity and direction can be controlled by the frequency of the piston that causes gas oscillations in the resonance tube. The obtained numerical solutions indicate that particle drift in a resonance tube enhances aerosol agglomeration. An agglomeration kernel is derived for this case, accounting for particle drift, leading to an estimate of agglomeration time. The time predicted by present model is of the same order of magnitude as that obtained from experiments in the literature. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1598968]

PACS numbers: 43.25.Gf, 43.25.Uv [MFH]

I. INTRODUCTION

Aerosol agglomeration is a process in which particles adhere to one another as a result of collisions forming agglomerates. Particle agglomeration may be used in air pollution control to enhance the performance of conventional particle filtration devices, such as scrubbers or cyclone separators, which are inefficient for small particles in the micrometer or submicrometer size range. For example, scrubber power requirements increase sharply with decreasing particle diameter below 0.5 μm.[1] A possible way to increase filtration efficiency is to increase the particle size by agglomeration and separate the agglomerates by a relatively inexpensive piece of equipment such as a cyclone separator.

Agglomeration of small particles by acoustic waves is a well-known phenomenon,[2,3] which has been investigated very intensively by many researchers. Yet, some effects are still not fully understood. The agglomeration in periodic shock waves arising in resonance tubes has been much less investigated, although the effect is more significant. Agglomeration of 1–10-μm aerosol particles has been investigated experimentally by Temkin.[4] He found that inside the resonance band, where the aerosol was influenced by periodic shock waves, the agglomeration was much more intensive than that outside the resonance band, where gas flow was continuous. Recently, Shuster *et al.*[5] have found vigorous agglomeration in periodic shock waves of a submicrometer aerosol, comprising of particles less than 0.3 μm in diameter.

Aerosol agglomeration is caused by the difference in the velocities of the agglomerating particles. In a resonance tube, such a difference may be caused by particle drift, which arises under the influence of periodic shock waves.[6] Since drift velocities strongly depend on the size of the particles, different sized particles drift with different velocities resulting in interparticle collisions. Hence, it is important to understand the physical mechanisms causing particle drift in a resonance tube, and to be able to calculate drift velocities. Once the drift velocities are obtained, an agglomeration kernel accounting for particle drift can be derived and aerosol agglomeration can be calculated.

The present paper considers motion of particles subjected to resonance gas oscillations produced by a sinusoidally moving piston at one end of a resonance tube. Experiments show that when the piston frequency is in a narrow band about the resonance frequency, shock waves appear in the tube[7] (Fig. 1). The shock waves travel back and forth along the tube with a speed close to that of sound. At that stage, the amplitude of gas pressure oscillations is amplified significantly compared to that outside the resonance band. The problem of resonance gas oscillations was treated analytically[8,9] by asymptotic expansions in terms of a small parameter $\varepsilon=\sqrt{\pi l/L}$, where l and L are the piston amplitude and tube length, respectively. It was found that in the resonance band, the disturbances in the thermodynamic variables are of order $O(\varepsilon)$, and the axial gas velocity of $O(\varepsilon C_0)$, while for nonresonance frequencies, the disturbances are of $O(\varepsilon^2)$, and the gas velocity of $O(\varepsilon^2 C_0)$, where C_0 is the speed of sound (see Fig. 1).

Longitudinal motion of small rigid particles in a resonance tube was treated analytically by Goldshtein *et al.*,[6] who found that nonsymmetrical periodic shock waves cause particle drift from the middle cross-section of the resonance tube toward its ends, as observed in experiments. To facilitate an analytical solution, they made several simplifying assumptions. Specifically, they assumed that the gas velocity in any cross-section of the tube is uniform between shocks and approximated by a stepwise function. Moreover, the change in gas density during the oscillations was neglected. Accordingly, they obtained that particle drift to the tube ends is caused by them lagging after the shock, with drift velocities strongly dependent on the ratio τ/T_{res}, where $T_{res}=2L/C_0$ is the period of resonance oscillations, and τ is the particle relaxation time. They found that particle drift is vigorous when $\tau/T_{res}\sim 1$ and negligible when $\tau/T_{res}\ll 1$. Thus, small particles, with relaxation times significantly shorter than the resonance oscillation period, oscillate around their equilib-

a)Author to whom correspondence should be addressed. Electronic mail: gutfinge@techunix.technion.ac.il

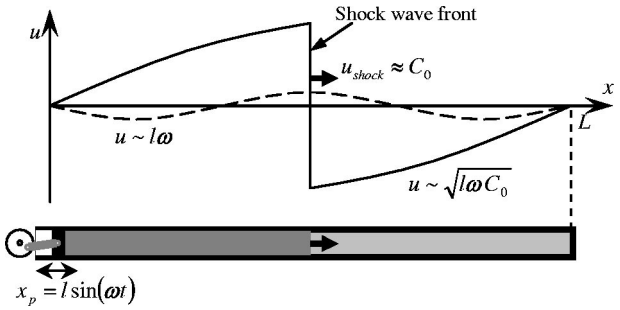


FIG. 1. Schematic of gas oscillations within a closed tube driven by an oscillating piston. The dashed line shows velocity distribution with the tube at a frequency f far from the resonance frequency $f_{res}=C_0/2L$, while the solid line shows velocity distribution at resonance, when $f=f_{res}$.

rium position, while large particles do drift toward the tube ends.

Goldshtein *et al.*[6] used their model to explain Temkin's experimental results.[4] They showed that there is a significant difference in oscillation amplitudes of 1- and 10-μm particles in the experiments of Temkin, and assumed that agglomeration was due to orthokinetic interaction[10] between the particles caused by this amplitude difference. Actually, the orthokinetic agglomeration mechanism should be supplemented by a refilling mechanism responsible for particles approaching to within a distance where orthokinetic interaction can occur.[3] In turbulent resonance oscillations,[11] for example, refilling may be caused by particle diffusion due to turbulent gas fluctuations. Without the refilling mechanism, the agglomeration stops after several oscillation periods, contrary to experimental observations. For example, in Temkin's experiment, performed in a tube characterized by laminar flow, the agglomeration was completed after about 100 s, which corresponds to thousands of oscillation periods. Hence, only with the help of refilling his results can be explained. Although the model of Goldshtein *et al.*[6] exhibits a difference in oscillation amplitude, it practically predicts zero drift for particles in the range 1–10 μm and cannot provide a plausible explanation for Temkin's experimental results.

In the present work, we attempt to take into account more details of gas oscillations at resonance by treating numerically longitudinal particle motion in a resonance tube. To this end, we calculate one-dimensional gas motion induced in the tube by an oscillating piston by means of a finite difference algorithm. One of our goals is to investigate the physical mechanisms of particle refilling during agglomeration in laminar flow. Moreover, through the paper we compare the predictions of the analytical model of Goldshtein *et al.*,[6] which was derived for a gas oscillating at the resonance frequency only, with our numerical solution. The present study also considers particle drift due to oscillations outside the resonance frequency.

The paper is organized as follows:

(i) In Sec. II, we present the equations of gas motion in a resonance tube, together with the initial and boundary conditions imposed on the gas.
(ii) The forces acting on an aerosol particle inside a resonance tube are introduced in Sec. III.
(iii) In Sec. IV, we study particle motion arising when the tube is driven at resonance, while
(iv) Section V presents the solution obtained for frequencies around resonance.
(v) In Sec. VI, we assess the effect of particle drift on aerosol agglomeration inside a resonance tube and derive a kernel for this case.
(vi) We summarize our results in the closure given in Sec. VII.

II. GOVERNING EQUATIONS

The basic equations governing one-dimensional isentropic flow of a perfect gas in a tube are those of continuity, momentum and state:

$$\rho_t+(\rho u)_x=0, \tag{1}$$

$$\rho(u_t+uu_x)+p_x=0, \tag{2}$$

$$p/p_0=(\rho/\rho_0)^{\gamma}, \tag{3}$$

where u is the axial gas velocity, t is time, γ is the adiabatic exponent, p is pressure and ρ is gas density. Here, the index 0 denotes the initial quantities.

The boundary conditions for the gas velocity are

$$u|_{x=x_p(t)}=\dot{x}_p(t), \qquad u|_{x=L}=0. \tag{4}$$

Here, x_p and $\dot{x}_p$ are the coordinate and velocity of the piston, respectively,

$$x_p(t)=l\sin(\omega t), \qquad \dot{x}_p(t)=l\omega\cos(\omega t), \tag{5}$$

and $\omega=2\pi f$ is the piston angular velocity.

Equations (1)–(3) with the boundary conditions (4) were solved numerically by an implicit finite difference algorithm.[12] This algorithm was successfully applied for many applications, including continuous flows and flows with shock waves.[13] We used a grid with a spatial step of $\Delta x=[L-x_p(t)]/600$, and a time marching step of $\Delta t=5\times10^{-5}f^{-1}$. The accuracy of the algorithm is of $O(\Delta x^2)$ in space, and $O(\Delta t)$ in time. To test the grid quality, its density was increased, indicating that an increase in grid density practically does not affect the accuracy of the solution. Initial conditions of the simulation were $u=0$, $p=p_0=10^5$ Pa and $\rho=\rho_0=1.2$ kg/m^3. The simulation was continued until a time periodic solution was obtained.

Recently, it was shown that the isentropic model of resonance gas oscillation is valid for small ε only.[14] In the case of $\varepsilon\geqslant0.1$, the flow is typically turbulent[11] and the isentropic model overestimates the experimentally measured pressure amplitude up to 45%. An isothermal numerical model was proposed[14] including some ad hoc terms that account for heat and viscous interactions between the gas and the tube. The isothermal model showed good agreement with experimental data for $\varepsilon\geqslant0.1$, however these ad hoc terms had to be adjusted based on experimental measurements of the heat transfer coefficient.

In spite of the differences in the motion of the gas, as calculated from the two models, when it comes to particle drift the difference between the results obtained from the isothermal and isentropic models for the same pressure am-

TABLE I. Drag coefficient correlations.

Range	Correlation
$\mathrm{Re}_p<0.01$	$C_D=\frac{3}{16}+\frac{24}{\mathrm{Re}_p}$
$0.01<\mathrm{Re}_p<20$	$C_D=\frac{24}{\mathrm{Re}_p}[1+0.1315\,\mathrm{Re}_p^{(0.82-0.05\log_{10}\mathrm{Re}_p)}]$
$20<\mathrm{Re}_p<260$	$C_D=\frac{24}{\mathrm{Re}_p}[1+0.1935\,\mathrm{Re}_p^{0.6035}]$

plitude is rather small. Hence, we conclude that the physical mechanisms governing particle drift remain the same for both models, and use in our calculations the simpler isentropic model. The latter model was also used in Ref. 6 to calculate particle drift.

III. FORCES ACTING ON A PARTICLE

The drag on bodies in unsteady motion in fluids has been the subject of many theoretical and experimental investigations.[15,16] The motion of a particle in a one-dimensional flow is affected by the forces due to Stokes drag, pressure gradient, added mass and that given by the Basset integral.

The influence of unsteadiness on the force acting on a particle exposed to an oscillating fluid is defined by the following dimensionless number $K=\omega d_p^2/\nu$,[17] where d_p is the particle diameter and ν is the kinematic viscosity. When $K\ll 1$, all the forces except the drag force may be neglected. For a typical resonance tube, where the piston frequency does not exceed 100 Hz, and an aerosol with micrometer-size particles, the condition $K\ll 1$ is satisfied. Moreover, for typical aerosols, in which the particle-gas density ratio is about $\rho_p/\rho\sim 10^3$, the effect of added mass and Basset force on particle trajectory is negligible.[2,15] Under these assumptions, the motion of small particles is governed by the Stokes force only.

Hence, the motion of a spherical particle with diameter d_p, density ρ_p and velocity u_p, in a fluid whose instantaneous velocity near the particle is u can be expressed as[15]

$$\frac{du_p}{dt}=\frac{3}{4}\frac{C_D}{d_p}\frac{\rho}{\rho_p}u_r|u_r|, \tag{6}$$

where $u_r=u-u_p$ is the relative fluid-particle velocity and C_D is the drag coefficient.

Correlations for the drag coefficient lift C_D for a spherical particle are given in Table I. These correlations correspond to the so-called standard curve derived from many experimental data.[15] It should be noted that the standard curve was obtained under uniform flow conditions. However, in many cases, the fluid flow itself is turbulent even when the relative velocity u_r and the resulting particle Reynolds number $\mathrm{Re}_p=d_p|u_r|/\nu$ are small. The local turbulent intensity of the fluid causes deviations in C_D from the standard curve. Unfortunately, no generally valid relations between the drag coefficient C_D and the local turbulent intensity of the fluid are available.[16] On the other hand, when the particle Reynolds number Re_p is low, i.e., the drag coefficient C_D is a linear function of the relative velocity, the average particle velocity is unaffected by turbulence[15] and the standard drag coefficients of Table I may be used.

When gas compressibility cannot be neglected, temperature fluctuations can be significant, affecting the dynamic viscosity μ. In that case, the viscosity can be calculated from Sutherland's law[18]

$$\mu=C_1\frac{\theta^{3/2}}{\theta+C_2}, \tag{7}$$

where θ is the absolute temperature, $C_1=1.458\times 10^{-6}$ kg/(m·s·K) and $C_2=110.4$ K for air at moderate temperatures.

To verify the above assumptions, we use Eq. (6) to calculate particle trajectories in comparison with an equation accounting for pressure gradient, added mass and Basset integral. We found that the trajectories were practically identical in both cases for particle diameters, in the range 0.2–100 μm. Thus, we conclude that the assumptions leading to Eq. (6) are valid and this equation may be used for the calculations of the motion of small particles in a resonance tube.

Equation (6) was solved by a standard predictor-corrector method[13] with a time marching step of $\Delta t=5\times 10^{-2}\tau$, where $\tau=1/18\,\rho_p d_p^2/\mu$. To test the numerical solution, the time step was decreased, indicating no effect on the accuracy of the solution. The particle density was set of $\rho_p=10^3\rho_0$.

IV. PARTICLE DRIFT AT RESONANCE FREQUENCY

Goldshtein *et al.*[6] found that nonsymmetrical periodic shock waves cause particle drift from the middle cross-section toward the ends of the resonance tube. It follows from their mathematical model that particles drift to the tube ends with different velocities, which are strongly dependent on the ratio τ/T_{res}. According to that model, small particles, with $\tau/T_{res}\ll 1$, oscillate around their equilibrium positions, while large particles with $\tau/T_{res}\sim 1$ drift toward the tube ends.

The analytical model of Goldshtein *et al.*[6] was verified experimentally, using a 3.5-mm spherical particle, and a qualitative agreement was found. It should be noted that particle drift measurement is a rather complex task. During its motion, the particle oscillates with a velocity of the order of the gas velocity and a frequency that equals that of the piston. For example, in the experiments of Goldshtein *et al.*,[6] the particle velocity was as high as 60 m/s and the resonance frequency was $f_{res}=1/T_{res}\sim 46.6$ Hz, while the drift velocity, according to the model, was of the order of 1 m/s, making its measurement rather difficult.

Figure 2 presents the dimensionless drift velocity, u_d, of particles with different diameters at the resonance frequency f_{res}, calculated numerically as well as analytically.[6] In this figure, data are presented for a resonance tube 365 cm long and with a 5.4-cm internal diameter, corresponding to that used in the experiments.[5,6] The piston amplitude was $l=2.2$ cm corresponding to $\varepsilon=0.137$.

Following the analytical model,[6] the drift velocity in Fig. 2 was set in dimensionless form by dividing it by

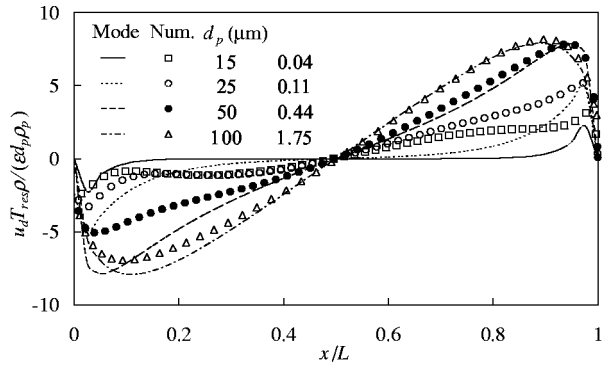


FIG. 2. Dimensionless particle drift velocity as a function of distance from the piston in a tube with $\varepsilon=0.137$ at $f_{res}=46.6$ Hz. The lines were calculated using the analytical model.[6] The points indicate the numerical results.

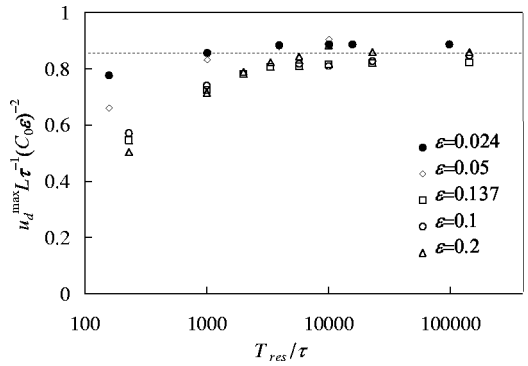


FIG. 4. Maximum values of the dimensionless drift velocity versus the ratio T_{res}/τ calculated numerically for tubes with different ε at the resonance frequency. The dashed line shows the value of the dimensionless maximum drift velocity $\tilde{u}_d^{max}=5/6$.

$\varepsilon d_p \rho_p / T_{res}\rho$. As seen, the model predictions are in fair agreement with the results of the numerical calculations for particles with $\tau/T_{res}\sim 1$. In Fig. 2, the maximal dimensionless drift velocity is obtained for a particle with $d_p=50$ μm, while for other particles the drift velocity is smaller. For particles with $d_p=25$ μm and $d_p=15$ μm, the model fits the numerical results near the tube ends and at the mid-cross-section $x/L=0.5$, while, away from the tube ends, it significantly underestimates the drift velocity.

According to the analytical model, a particle drifts because it lags after the shock wave. The shock wave travels along the tube and passes through the particle twice during every oscillation period. When a particle with $\tau/T_{res}\sim 1$ is not at the middle cross section, $x/L=0.5$, the times between the shock passing back and forth are unequal, leading to asymmetry and particle drift. The analytical model predicts that particles with a small ratio $\tau/T_{res}\ll 1$ follow the gas perfectly, and, hence, do not drift. On the other extreme, large particles characterized by $\tau/T_{res}>1$ do not follow the gas oscillations and remain near their initial positions. Hence, drift is vigorous for particles having $\tau/T_{res}\sim 1$ and decreases for particles with larger and smaller τ/T_{res} ratios.

Figure 3 presents dimensionless drift velocities, $u_d L\tau^{-1}(C_0\varepsilon)^{-2}$, calculated for particles with $\tau/T_{res}\ll 1$ and a resonance frequency f_{res}, in a tube with $\varepsilon=0.137$. We put the drift velocity u_d in dimensionless form by dividing it by

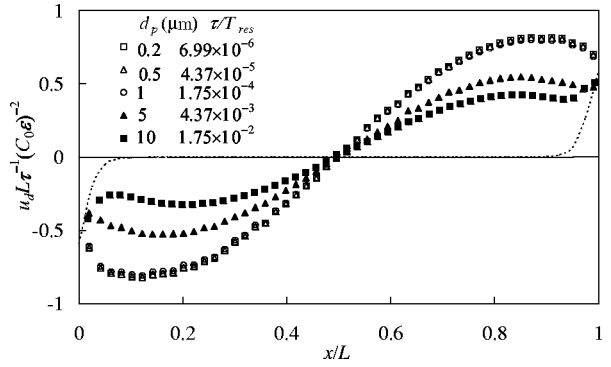


FIG. 3. Dimensionless particle drift velocity as a function of distance from the piston in a tube with $\varepsilon=0.137$ at $f_{res}=46.6$ Hz for $\tau/T_{res}\ll 1$. The dotted and solid lines were calculated using the analytical model[6] for $d_p=10$ μm and $d_p=5$ μm, respectively. The points indicate the numerical results.

the expression obtained in Eq. (13) given below. This differs from the dimensionless velocity of Fig. 2, which was applicable for rather large particles with $\tau/T_{res}\sim 1$, and is not applicable here. One can see that here the drift velocities depend weakly on particle diameter. Moreover, when particle diameter decreases, such that $\tau/T_{res}<2\times10^{-4}$, the dimensionless velocities converge into a common curve. We compare our results with the predictions of the analytical model.[6] It can be seen that the predictions of the analytical model are inadequate for particles with $\tau/T_{res}\ll 1$. This can be explained by the fact that the model of Goldshtein *et al.* ignores the velocity and density variations of the gas between the shocks, assuming them to be constant, while the numerical model intrinsically accounts for the variations that govern the drift of particles with $\tau/T_{res}\ll 1$, as discussed below. On the other hand, in the case of $\tau/T_{res}\sim 1$ when the drift is governed by shocks, agreement between the models is good.

Figure 4 shows dimensionless values of the maximum drift velocity $u_d^{max} L\tau^{-1}(C_0\varepsilon)^{-2}$ as a function of T_{res}/τ for tubes with different ε at resonance frequency. It is remarkable that when T_{res}/τ increases, the dimensionless maximum velocities converge to the same value of about $\tilde{u}_d^{max}\cong 5/6$, independent of the tube parameter ε. The critical value of T_{res}/τ beyond which the velocity becomes independent of ε and T_{res}/τ is about 5×10^3. For $T_{res}/\tau>5\times10^3$, the deviation from the value of 5/6 does not exceed 5% and is of the order of the numerical error.

Figure 5 presents the dimensionless drift velocities for particles with $\tau/T_{res}=10^{-4}$, calculated for tubes with different ε. As seen, the plots of the dimensionless drift velocities collapse into a single curve. Hence, the dimensionless drift velocity is independent not only of particle diameters as shown in Fig. 3, but also of tube parameters when $\tau/T_{res}\ll 1$. Thus, the curve is a universal one.

In order to explain the drift of particles with $\tau/T_{res}\ll 1$, we consider the motion of a 0.5-μm particle with $\tau/T_{res}=4.4\times10^{-5}$ initially placed at $x=0.25L$ inside a tube with $\varepsilon=0.137$. Figure 6 presents the time evolution of the velocity of the gas, u, surrounding this particle and the relative fluid-particle velocity, $u_r=u-u_p$. Even for very low relaxation times τ, the relative velocity u_r is different from zero. This is caused by the continuously accelerating or de-

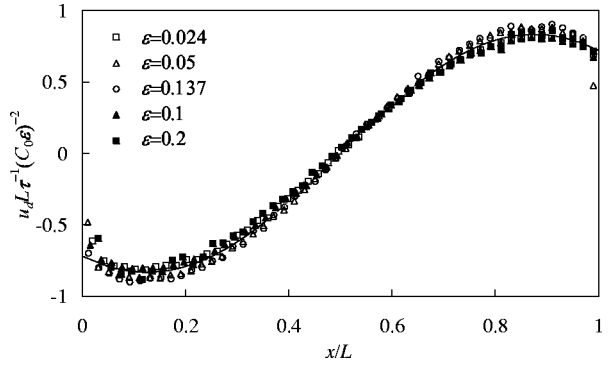


FIG. 5. Dimensionless particle drift velocity of a particle with $\tau/T_{res}=10^{-4}$ as a function of distance from the piston in tubes with different ε at the resonance frequency. The solid line is calculated by Eq. (14).

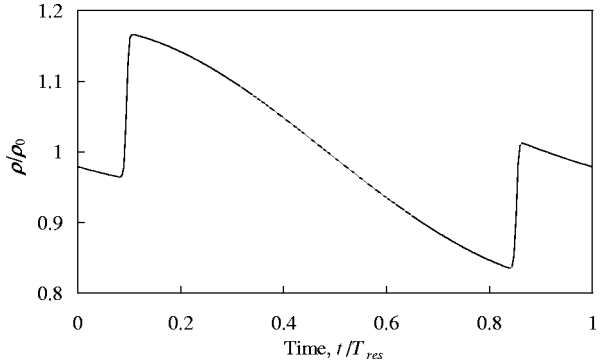


FIG. 7. Time evolution of the density of the gas surrounding a 0.5-μm particle ($\tau/T_{res}=4.4\times10^{-5}$) initially placed at $0.25L$ inside a tube with $\varepsilon=0.137$ at the resonance frequency.

celerating gas (Fig. 6). We can estimate the relative velocity u_r as follows. The relative velocity is very small at all times except for the short periods, when the shock passes thorough the particle. Hence, $\mathrm{Re}_p \ll 1$, and we may calculate the Stokes drag force as[16]

$$F_d = 3\pi\mu d_p u_r . \tag{8}$$

Since the gas velocity changes more rapidly than the relative velocity (Fig. 6), we assume that the gas and particle accelerations are approximately equal, leading to

$$\frac{du}{dt} = \frac{F_d}{m_p}, \tag{9}$$

where $m_p = 1/6\,\pi d^3 \rho_p$ is the mass of the particle. Substitution of Eq. (8) into Eq. (9) yields

$$u_r = \frac{d_p^2 \rho_p}{18\mu}\frac{du}{dt} = \tau\frac{du}{dt}. \tag{10}$$

The gas acceleration can be estimated as $du/dt \sim U_0\omega$,[9] where $U_0 \sim \varepsilon C_0$ is the characteristic gas velocity. Taking into account that at resonance the angular velocity is $\omega = 2\pi/T_{res} = \pi C_0/L$, we obtain $du/dt \sim \varepsilon C_0^2/L$. Substituting this expression into Eq. (10), we estimate the particle relative velocity as

$$u_r \sim \tau\frac{\varepsilon C_0^2}{L}. \tag{11}$$

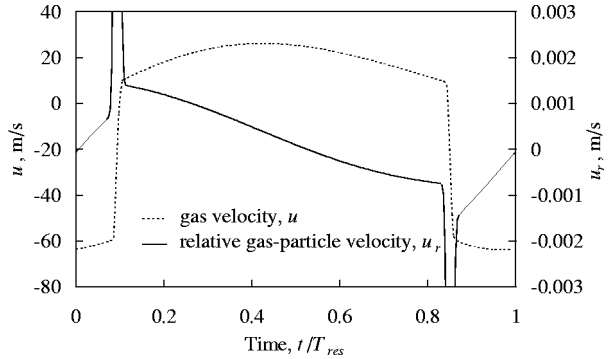


FIG. 6. Time evolution of the velocity u of the gas surrounding a particle and that of the relative gas-particle velocity, $u_r = u - u_p$. The particle diameter is 0.5 μm ($\tau/T_{res}=4.4\times10^{-5}$) and the initial position $0.25L$ inside a tube with $\varepsilon=0.137$ at the resonance frequency.

Figure 6 shows that the relative velocity u_r can be either positive or negative, with the average velocity being approximately zero. However, due to gas compressibility, its density changes along the particle trajectory, as shown in Fig. 7, where the density of the gas surrounding the particle is presented as a function of time. Comparing Figs. 6 and 7, we note that when the relative velocity is positive, the particle travels in a high density gas and *vice versa*. Hence, the particle does indeed drift within the gas and the drift velocity, u_d, is about

$$u_d \sim u_r\frac{\Delta\rho}{\rho_0}, \tag{12}$$

where ρ_0 is the initial gas density. Taking into account that $\Delta\rho/\rho_0 \sim \varepsilon$,[9] and combining Eq. (12) with Eq. (11), we obtain

$$u_d \sim \tau\frac{\varepsilon^2 C_0^2}{L}. \tag{13}$$

This expression is used to put the drift velocity for $\tau/T_{res} \ll 1$ into dimensionless form. The fact that the dimensionless drift velocities presented in Figs. 3 and 5 collapse into a single curve for particles with $\tau/T_{res} \ll 1$ supports our assumptions regarding the motion of such particles.

Using the scaling given by Eq. (13), we approximate drift velocity of particles with $\tau/T_{res} \ll 1$ as

$$u_d = -\tilde{u}_d^{\max}\tau\frac{\varepsilon^2 C_0^2}{L}\cos\left(\frac{\pi}{3}\left[4\frac{x}{L}-\frac{1}{2}\right]\right), \tag{14}$$

where $\tilde{u}_d^{\max} = 5/6$ is taken from Fig. 4 and the constants are chosen to fit the numerical data. The drift velocity calculated from Eq. (14) is shown by the solid line in Fig. 5 and is in excellent agreement with the numerical data except at the vicinity of the tube ends, where the drift velocity rapidly approaches zero. Since we use the maximal value of the dimensionless drift velocity in Eq. (14), the equation gives an upper limit of the drift velocity for particles with $\tau/T_{res} \ll 1$.

Based on the above assumptions for the case of $\tau/T_{res} \ll 1$, a model for particle drift can be derived from the model of resonance gas oscillation.[9] However, it was found that such a model underestimates the drift velocity at about 50% and includes some transcendental equations, hence making the analysis rather complicated. We attribute this discrepancy

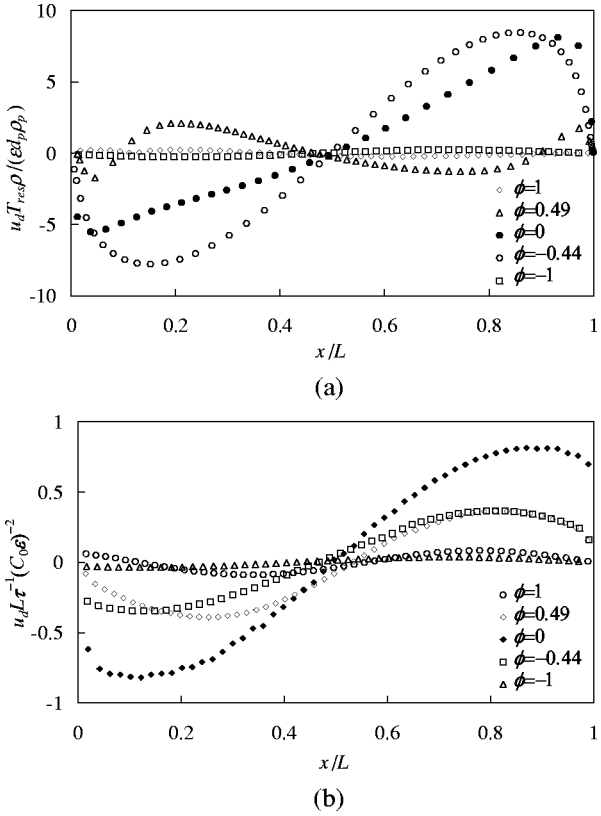


FIG. 8. Dimensionless drift velocity of a particle inside a tube with $\varepsilon=0.137$ as a function of position along the tube at different oscillation frequencies ϕ: (a) $d_p=50$ μm; (b) $d_p=0.5$ μm.

to the asymptotic nature of the model used to calculate gas oscillations.[9] Thus, the ad hoc model given by Eq. (14) seems to be more useful for practical applications than that derived from the model of resonance gas oscillation.

We distinguish between two types of particle drift mechanisms. The first is relevant for particles with $\tau/T_{res}\sim 1$, where the drift is governed by shock waves.[6] Here, the drift velocity is proportional to ε. The second mechanism, where the drift is due to gas acceleration and compressibility, is most pronounced for particles with $\tau/T_{res}\ll 1$. In this case, the drift velocity is of the order of ε^2. This case was not treated before. For particles with intermediate τ/T_{res}, both mechanisms hold. For example, the drift velocity of a 10-μm particle presented in Fig. 3 fits the model of Goldshtein *et al.*[6] near the tube ends, while in the middle of the tube the drift velocity is close to that for $\tau/T_{res}\ll 1$. Thus, in the vicinity of the tube ends, the drift of 10-μm particles is governed by shock waves, whereas in the middle of the tube, it is due to gas acceleration and compression.

V. PARTICLE DRIFT OUTSIDE THE RESONANCE FREQUENCY

Figures 8(a) and (b) present the dimensionless drift velocities of a 50-μm particle and a 0.5-μm particle, respectively, in a tube with $\varepsilon=0.137$ at different oscillation frequencies. Following Chester,[8] we define a dimensionless frequency ϕ relative to the resonance frequency f_{res} as $\phi=(f-f_{res})/\varepsilon f$. Thus, $\phi=0$ corresponds to the resonance frequency. Simulations show that in terms of ϕ, the resonance band, in which gas flow is accompanied by shocks, corresponds to the range $-0.56\leqslant\phi\leqslant 0.56$. As seen in Fig. 8, the drift velocity strongly depends on ϕ. When ϕ is outside of the resonance band, where the flow is continuous, the drift velocity is close to zero. The drift velocity achieves its maximum inside the resonance band, where the flow is discontinuous due to shocks. The shocks are strongest at the resonance frequency $\phi=0$, and weaker outside it. It is surprising that for $d_p=50$ μm [Fig. 8(a)] and a frequency $\phi=-0.44$, which is close to the lower bound of the resonance band, the drift velocity exceeds that at the resonance frequency $\phi=0$. Moreover, when the frequency is close to the upper bound of the resonance band, the drift velocity of a 50-μm particle changes its direction. Thus, for $\phi=0.49$, a 50-μm particle travels toward the middle of the tube, except at the vicinity of the tube ends, while at $\phi=0$, the same particle drifts toward the tube ends from any point in the tube. In contrast to the case of the 50-μm particle, the drift direction of a 0.5-μm particle is independent of the oscillation frequency [Fig. 8(b)]. Besides, when $d_p=0.5$ μm, the maximum drift velocity is registered at the resonance frequency $\phi=0$.

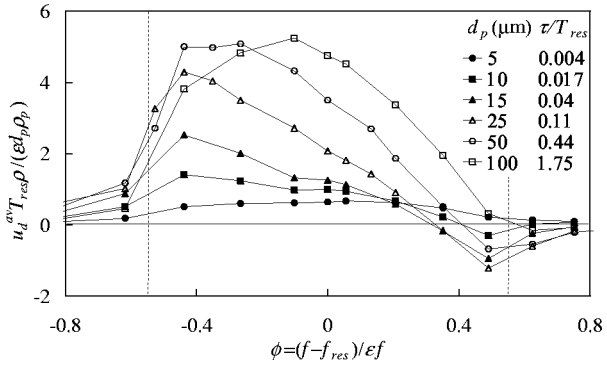


FIG. 9. Dimensionless average drift velocity for particles with different τ/T_{res} inside a tube with $\varepsilon=0.137$ versus dimensionless oscillation frequency ϕ.

In order to characterize the overall drift of the particles in a resonance tube at different oscillation frequencies, we introduce a spatial average drift velocity u_d^{av}, which is calculated as

$$u_d^{av}=\frac{1}{L}\int_0^L u_d^e dx, \tag{15}$$

where $u_d^e=-u_d$, for $x/L<0.5$, and $u_d^e=u_d$ everywhere else. Hence, the drift velocity toward the nearest tube end is positive, while the drift toward the tube middle is negative.

Figures 9 and 10 show the average drift velocity u_d^{av} versus dimensionless frequency ϕ for particles having $\tau/T_{res}\sim 1$ and $\tau/T_{res}\ll 1$, respectively.

In the case of $\tau/T_{res}\sim 1$ shown in Fig. 9, when calculation is performed for a tube with $\varepsilon=0.137$, u_d^{av} varies considerably for different τ/T_{res}, but some common features can still be depicted. The maximum values of u_d^{av} lie inside the resonance band, where $\phi<0$. The maximum of u_d^{av} is in the vicinity of the lower bound of the resonance band, when τ/T_{res} is in the range 0.02–0.5. For example, for $\tau/T_{res}\approx 0.11$, the maximum u_d^{av} corresponds to $\phi_{\max}\approx -0.44$ and

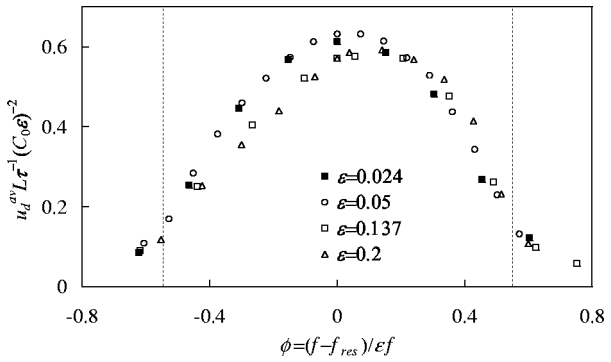


FIG. 10. Dimensionless average drift velocity of a 0.5-μm particle, for which $\tau/T_{res} \ll 1$, for tubes with different ε versus dimensionless oscillation frequency ϕ.

exceeds more than twice u_d^{av} at the resonance frequency $\phi = 0$. For larger and smaller τ/T_{res}, the ratio between maximum u_d^{av} and u_d^{av} for $\phi = 0$ decreases and $\phi_{\max}$ approaches the resonance frequency. Moreover, a negative drift, i.e., in the direction of the middle cross-section, is found in the vicinity of the upper bound of the resonance band. The maximum negative drift velocity is obtained for $\tau/T_{res} \approx 0.11$. Its value is about 3.5 times less than the maximum positive drift velocity of this particle. Note that for $\tau/T_{res} \approx 0.11$, the drift velocity is practically a linear function of ϕ in the range between the maximum positive and negative values of u_d^{av}. Hence, for particles with $\tau/T_{res} \sim 1$, the speed of drift can be readily controlled by frequency tuning, which is especially effective in the case of particles with $\tau/T_{res} \approx 0.1$.

For particles with $\tau/T_{res} \ll 1$ shown in Fig. 10, the drift velocities for different τ/T_{res} and ε converge to a common curve. The most vigorous drift prevails around $\phi = 0$. Outside the resonance band, the drift velocity decays to zero. In contrast to the case $\tau/T_{res} \sim 1$, the particles with $\tau/T_{res} \ll 1$ drift toward the tube ends only. Hence, for that case, the drift velocity can only be controlled by the oscillation frequency.

We, therefore, conclude that frequency tuning allows us to control particle drift. In the case of drift toward the tube ends, the drift velocity can be changed from zero to its maximum value for any particle. Moreover, particles having $\tau/T_{res} \approx 0.1$ can also drift toward the tube middle cross-section, when the piston is driven with an appropriate frequency. Still, the drift toward the middle cross-section is several times slower than that toward the tube ends. The maximum drift velocity toward the tube ends is at a frequency that is close to the lower bound of the resonance band, while that toward the tube middle occurs at a frequency near the upper bound.

VI. APPLICATION TO AEROSOL AGGLOMERATION

Temkin[4] showed that resonance gas oscillations cause rapid agglomeration of aerosol particles. In his experiments, he used a resonance tube with $\varepsilon = 0.024$ and $f_{res} = 68$ Hz, and particles with diameters in the range 1–10 μm with an initial concentration $N_0 \approx 10^4$ cm^{-3}. Goldshtein *et al.*[6] have estimated the oscillation amplitudes for the particles used by Temkin.[4] They found that the difference in the amplitudes between 1- and 10-μm particles is about 0.3 cm and exceeds significantly the initial distance between the particles, which are of the order of $\lambda \approx N_0^{-1/3} = 0.05$ cm. Based on this finding, it was suggested[6] that the agglomeration in Temkin's resonance tube is due to the orthokinetic mechanism.[10]

A particle in an oscillating fluid moves over a closed volume, termed the particle agglomeration volume.[3] Orthokinetic agglomeration takes place, when particle agglomeration volumes intersect due to the difference in particle oscillation amplitudes. However, during several oscillating periods, all the particles with crossing agglomeration volumes stick together, and no further agglomeration can occur until new particles are brought into these volumes by a refilling mechanism. Hence, the fact that particles oscillate with different amplitudes is insufficient for aerosol agglomeration in a resonance tube. Particle drift due to gas acceleration and compression is a good candidate for explaining such a refilling mechanism. Indeed, using Eq. (13), we can estimate that the drift velocity of a 10-μm particle is about 1 cm/s, while a 1-μm particle drifts at about 0.01 cm/s. Thus, 10-μm particles drift relative to 1-μm particles and refill the agglomeration volumes, leading to vigorous aerosol agglomeration, as observed by Temkin.[4]

Following Fuchs,[2] we define for drift in a resonance tube an agglomeration kernel for two particles with diameters d_{p1} and d_{p2}, such that $d_{p1} > d_{p2}$:

$$K(d_{p1}, d_{p2}) = \frac{\pi}{4} \hat{\zeta} d_{p1}^2 u_{12}, \tag{16}$$

where $u_{12} = |u_{d1} - u_{d2}|$ is the relative velocity of the particles and $\hat{\zeta}$ is the total capture coefficient. When two particles move with different velocities, there are two basic mechanisms that can bring them to touch each other: interception and Brownian diffusion. The capture coefficient of a single sphere corresponding to the interception mechanism is given by[2]

$$\zeta_i = \left(\frac{d_{p2}}{d_{p1}} + 1 \right)^2 - \frac{3}{2} \left(\frac{d_{p2}}{d_{p1}} + 1 \right) + \frac{0.5}{1 + d_{p2}/d_{p1}}. \tag{17}$$

The capture coefficient due to Brownian diffusion depends on the Peclet number,[2] $\mathrm{Pe} = d_{p1} u_{12} / D_2$. Here, D_2 is the mass diffusion coefficient of the smaller particle. For the case of $\mathrm{Pe} \gg 1$, which is characteristic of micrometer sized particles oscillating in a resonance tube, the capture coefficient can be calculated as[19]

$$\zeta_B \approx 4 \mathrm{Pe}^{-2/3}. \tag{18}$$

When $d_{p1} \gg d_{p2}$, leading to $\zeta_i \ll 1$ and $\zeta_B \ll 1$, the capture coefficient due to interception and diffusion may be approximated with good accuracy by $\zeta = \zeta_i + \zeta_B$.[20]

The change of concentration of the smaller particles within an agglomeration volume can be estimated as

$$-\frac{dn}{dP} = \zeta n, \tag{19}$$

where P is the number of oscillation periods, and n is the number concentration of particles having d_{p2}. The solution of Eq. (19) is $n = n_0 \exp(-\zeta P_d)$, where n_0 is the initial particle concentration and $P_d = \Delta \ell f_{res} / u_{12}$ is the number of pe-

riods during which the larger particle drifts through the agglomeration volume of length $\Delta\ell$. Defining the total capture coefficient as $\hat{\zeta}=(n_0-n)/n_0$, we obtain

$$\hat{\zeta}=1-\exp(-\zeta P_d). \tag{20}$$

The agglomeration volume length $\Delta\ell$ is equal to the difference between the oscillation amplitudes of the particles. In a resonance tube, the shock wave has a velocity jump of about εC_0. Due to their inertia, particles lag after the shock, and the particle oscillation amplitude is less than that of the gas by $\tau\varepsilon C_0$. Hence, we estimate the agglomeration volume length as $\Delta\ell=2\varepsilon C_0(\tau_1-\tau_2)$. Taking into account Eq. (14), we therefore obtain for particles with $\tau/T_{res}\ll 1$

$$P_d=\left[\varepsilon\tilde{u}_d^{\max}\left|\cos\left(\frac{\pi}{3}\left[4\frac{x}{L}-\frac{1}{2}\right]\right)\right|\right]^{-1}. \tag{21}$$

We should note that P_d depends on the tube parameter ε and particle coordinate only, and its minimal value is $P_d^{\min}=[\varepsilon\tilde{u}_d^{\max}]^{-1}$. Substituting this expression into Eq. (20), we obtain the minimal value of the total capture coefficient as

$$\hat{\zeta}_{\min}=1-\exp(-\zeta/[\varepsilon\tilde{u}_d^{\max}]). \tag{22}$$

Using Temkin's experimental conditions and particles with $d_{p1}=10$ μm and $d_{p2}=1$ μm, one obtains $\zeta_i\approx 0.0145$ and $\zeta_B\approx 0.0164$ leading to $\hat{\zeta}_{\min}\approx 0.8$.

Taking into account that for particles having $\tau/T_{res}\ll 1$ the drift velocity is given by Eq. (14), we rewrite Eq. (16) as

$$K(d_{p1},d_{p2})=\hat{\zeta}\frac{\pi}{4L}\tilde{u}_d^{\max}\varepsilon^2C_0^2d_{p1}^2 \times\left|(\tau_1-\tau_2)\cos\left(\frac{\pi}{3}\left[4\frac{x}{L}-\frac{1}{2}\right]\right)\right|. \tag{23}$$

Equation (23) can be substituted into the agglomeration equation[2] to calculate aerosol evolution in a resonance tube. However, aerosol concentration changes not only by agglomeration but also due to particle deposition caused by drift. Moreover, at resonance the motion of particles may be affected by secondary flow.[11] Hence, to simulate aerosol evolution at resonance, one should solve a set of aerosol conservation equations[2] for different sized particles within the two-dimensional flow field. These equations have to be supplemented by the agglomeration equation accounting for particle growth.

This is a rather formidable task. Instead of doing this, we estimate the time needed for aerosol agglomeration to be completed, when caused by particle drift only. For the sake of simplicity, we consider $u_{d1}\gg u_{d2}$ and neglect the dependence on the x-coordinate, leading to

$$K=\hat{\zeta}\frac{\pi}{4L}\tilde{u}_d^{\max}\varepsilon^2C_0^2d_{p1}^2\tau_1. \tag{24}$$

We note that the condition $u_{d1}\gg u_{d2}$ can be easy satisfied, since $u_d\sim d_p^2$.

The kernel K represents the volume through which a single particle with d_{p1} passes per unit time, capturing smaller particles. Hence, considering that the number concentration of particles with d_{p1} is of order of N_0, we calculate a dimensionless volume V covered by all such particles during time t as

$$V\sim N_0Kt. \tag{25}$$

Moreover, when $V=1$, all the particles in the tube are captured by the particles with d_{p1} and agglomeration process is complete. Hence, we estimate the time t_{agg} required to complete agglomeration as

$$t_{agg}\sim\frac{1}{N_0K}=\frac{1}{N_0}\frac{72L\mu}{\hat{\zeta}\pi\tilde{u}_d^{\max}\rho_p\varepsilon^2C_0^2d_{p1}^4}. \tag{26}$$

Substituting into Eq. (26) the parameters of Temkin's experiment, we obtain that $t_{\min}\sim 190$ s. This result is close to that of Temkin,[4] who found that after about 100 s the aerosol has been dissipated and the agglomeration process practically completed. We also note that that time is enough for 10-μm particles to migrate a distance of half a tube and reach the tube end. During their migration, these particles collide with smaller particles on their way, producing agglomerates having larger drift velocities than the initial particles. These agglomerates continue to travel and collide with other small particles until the end of the resonance tube is reached. Hence, we require that the time of migration should actually be longer than t_{agg} for the agglomeration to be completed.

More recently, Shuster *et al.*[5] have found intensive agglomeration of an initially submicrometer aerosol in a 365-cm-long resonance tube with $\varepsilon=0.137$. Aerosol concentration was $N_0\approx 10^8$ cm^{-3} leading to $\lambda\approx 22$ μm and the largest particles in the initial aerosol were 0.3 μm in diameter. After a 30-s experiment, the authors have registered particles in the range 1–3 μm as a result of the agglomeration process and, after about 50 s, particle agglomeration was complete. According to Eq. (14), the drift velocity of 0.3-μm particles is about 0.017 cm/s. Using Eq. (26), we obtain that the time of agglomeration for these particles should be larger than 800 s, which is higher by an order of magnitude than that found by experiments. The flow in a resonance tube with $\varepsilon\geqslant 0.1$ is typically turbulent and accompanied by secondary flows.[11] Hence, we can hypothesize that turbulent diffusion, which is most pronounced for submicrometer particles,[2] leads to particle collisions at the initial stage of the agglomeration process. After some growth of the particles, the time needed for agglomeration, which is proportional to d_p^{-4} [see Eq. (26)], rapidly decreases reducing the overall time of agglomeration. This hypothesis requires additional verification.

VII. CLOSURE

In this work, we investigated numerically the drift of small aerosol particles in a resonance tube and its effect on particle agglomeration. When a gas trapped inside a tube oscillates at its resonance frequency, particles travel toward the tube ends. We found that different physical mechanisms define particle drift depending on the ratio τ/T_{res}, where T_{res} is the period of resonance oscillations and τ is the particle relaxation time. For particles with $\tau/T_{res}\sim 1$, drift is caused by periodic shock waves traveling back and forth along the tube, while for those with $\tau/T_{res}\ll 1$ drift occurs as

a result of gas acceleration and compression. For particles with intermediate τ/T_{res}, both mechanisms affect particle motion.

For the case of $\tau/T_{res}\sim 1$, the results of the numerical model are in agreement with previous analytical solution of Goldshtein *et al.*[6] In this case, the drift velocity is proportional to $\varepsilon=\sqrt{\pi l/L}$, where l is the piston amplitude and L is the tube length. For particles having $\tau/T_{res}\ll 1$, we showed that the drift velocity is of order ε^2, and proposed a simple expression for velocity calculation at the resonance frequency, given by Eq. (14).

We also found that particle drift can be controlled by piston frequency. When the frequency is outside the resonance band, the drift is negligible and all the particles oscillate around their equilibrium positions. Within the resonance band, particles with $\tau/T_{res}\ll 1$ drift toward tube ends and their drift velocity is highest around the resonance frequency. For particles with $\tau/T_{res}\sim 1$, the maximum drift velocity is at a frequency that is close to that of the resonance lower bound and the drift is directed toward the tube ends. At frequencies near the upper bound, the drift is in the opposite direction, toward the middle cross-section of the tube. Hence, for these particles, either drift direction can be obtained by tuning of the piston frequency.

The effect of particle drift on aerosol agglomeration in a resonance tube was demonstrated. It was shown that the agglomeration process is initiated by the refilling mechanism caused by particle drift, and proceeds by orthokinetic particle interactions. We derived an expression for the agglomeration kernel accounting for drift of particles with $\tau/T_{res}\ll 1$. Using this kernel, we calculated the time required to complete the agglomeration process in a resonance tube and compared it with data from the literature. It was found that the kernel leads to an agglomeration time of the same order of magnitude as that obtained experimentally for a tube with laminar gas flow. However, in the case of turbulent gas motion, the calculated agglomeration time exceeds significantly that obtained in experiments, necessitating to account for turbulence as the cause of refilling to describe the agglomeration process in that case.

ACKNOWLEDGMENT

This research was supported by the Fund for the Promotion of Research at the Technion, by the Israel Science Foundation.

[1] D. S. Scott, "A new approach to the acoustic conditioning of industrial aerosol emissions," J. Sound Vib. **43**, 607–619 (1975).
[2] N. A. Fuchs, *The Mechanics of Aerosols* (Pergamon, Oxford, 1964).
[3] E. P. Mednikov, *Acoustic Coagulation and Precipitation of Aerosols* (Consultants Bureau, New York, 1965).
[4] S. Temkin, "Droplet agglomeration in a shock wave flow field," Phys. Fluids **13**, 1639–1641 (1970).
[5] K. Shuster, M. Fichman, A. Goldshtein, and C. Gutfinger, "Agglomeration of submicrometer particles in weak periodic shock waves," Phys. Fluids **14**, 1802–1805 (2002).
[6] A. Goldshtein, K. Shuster, P. Vanishtein, M. Fichman, and C. Gutfinger, "Particle motion in resonance tubes," J. Fluid Mech. **360**, 1–20 (1998).
[7] R. A. Saenger and G. E. Hudson, "Periodic shock waves in resonating gas columns," J. Acoust. Soc. Am. **32**, 961–971 (1960).
[8] W. Chester, "Resonant oscillations in closed tubes," J. Fluid Mech. **18**, 44–64 (1964).
[9] A. Goldshtein, P. Vainshtein, M. Fichman, and C. Gutfinger, "Resonance gas oscillations in closed tubes," J. Fluid Mech. **322**, 147–163 (1996).
[10] O. Brand, H. Freund, and E. Hiedemann, "Schwebstoffe im Schallefeld," Z. Phys. **104**, 511–533 (1937).
[11] P. Merkli and H. Thomann, "Transition to turbulence in oscillating pipe flow," J. Fluid Mech. **68**, 567–575 (1975).
[12] R. M. Beam and R. F. Warming, "An implicit finite-difference algorithm for hyperbolic systems in conservation law form," J. Comput. Phys. **22**, 87–110 (1976).
[13] D. A. Anderson, J. C. Tannehill, and R. H. Pletcher, *Computational Fluid Mechanics and Heat Transfer* (Hemisphere, New York, 1984).
[14] A. Alexeev, A. Goldshtein, and C. Gutfinger, "Heat interaction in a resonance tube," Phys. Fluids **14**, 1812–1815 (2002).
[15] R. Clift, J. R. Grace, and M. E. Weber, *Bubbles, Drops and Particles* (Academic, New York, 1978).
[16] L.-S. Fan and C. Zhu, *Principles of Gas-Solid Flow* (Cambridge U.P., Cambridge, 1998).
[17] R. I. Nigmatulin, *Dynamics of Multiphase Media, Annals of Nuclear Energy* (Hemisphere, Washington, 1990).
[18] H. Schlichting, *Boundary Layer Theory* (McGraw–Hill, New York, 1979).
[19] V. G. Levich, *Physico-Chemical Hydrodynamics* (Prentice–Hall, New York, 1962).
[20] C. Gutfinger and G. Tardos, "Theoretical and experimental investigation on granular bed dust filters," Atmos. Environ. **13**, 853–867 (1979).

Properties of the zeroth-, first-, and higher-order approximations of attributes of elastic waves in weakly anisotropic media

Véronique Farra[a] and Ivan Pšenčík[b]
Département de Sismologie, Institut de Physique du Globe de Paris, 4 Place Jussieu, 75252 Paris Cedex 05, France and Geophysical Institute, Academy of Sciences of the Czech Republic, Boční II, Praha 4, Czech Republic[c]

(Received 1 November 2002; revised 28 April 2003; accepted 15 May 2003)

Use of the perturbation theory in the study of attributes of elastic waves propagating in weakly anisotropic media leads to approximate but transparent and simple formulas, which have many applications in forward and inverse wave modeling. We present and study such formulas. We show that all studied attributes depend on elements of a matrix linearly dependent on parameters of a medium. We study this dependence with the goal to understand which parameters of the medium, and in which combinations, affect individual wave attributes. Alternative auxiliar vector bases, in which the matrix can be specified, are proposed and studied. The vector bases offer alternative specifications of polarization vectors of qS waves. One of the important observations is that the higher-order ($n \geqslant 2$) perturbation formulas for qS waves are obtained separately for $qS1$ and $qS2$ waves. We also study effects of the use of the perturbation theory on the accuracy of the determination of the acoustical axes in weakly anisotropic media. We show that longitudinal directions in the first-order approximation are identical with actual ones. In singular directions, however, the first-order formulas provide directions, which may deviate from the exact ones, or they may even indicate false singular directions. Again, the above-mentioned matrix depending linearly on the parameters of the medium plays a central role in this study. [DOI: 10.1121/1.1591772]

PACS numbers: 43.25.Lj, 43.25.Zx [DEC]

I. INTRODUCTION

Perturbation theory is a useful tool for approximate study of wave properties in weakly anisotropic media. Especially its first-order approximation yields simple and transparent formulas only slightly more complicated than those for isotropic media. Thus it allows approximate but very effective solution of forward and inverse problems like ray tracing, travel-time computations, travel-time tomography, local determination of parameters of media, etc. Perturbation theory also plays an important role in the coupling ray theory and quasi-isotropic approximation in particular [see Červený (2001), for example].

Farra (2001) extended standard techniques for the calculation of the first-order approximations of the phase velocity and polarization vectors (see, e.g., Jech and Pšenčík, 1989), and proposed a procedure how to calculate even higher-order approximations. In this contribution, we extend and apply results of Farra (2001). We extend it by a detailed description of the construction of perturbation series for qS waves, by introduction of new, useful, auxiliary vectorial bases, in which the perturbation formulas can be specified, and by the sensitivity study of the wave attributes. We then apply the formulas to study, both analytically and numerically, behavior of first-order perturbation formulas in the directions of acoustical axes.

In perturbation formulas of any order, an important role is played by a matrix, B_{mn}, whose elements control various attributes of elastic waves. The matrix has several interesting and important properties. It is independent of the choice of a reference medium. It depends linearly on weak anisotropy (WA) parameters, which are, in turn, linearly related to elastic parameters. Explicit dependence of the matrix B_{mn} on various parameters of the medium or their combinations makes it possible to study sensitivity of various attributes of elastic waves propagating in an anisotropic medium to its parameters. This is important information for inversion of observed data into elastic moduli (Chu *et al.*, 1994; Song *et al.*, 2001; Zheng and Pšenčík, 2002). Attention is also paid to alternative ways of constructing approximate formulas for the polarization vectors of qS waves.

A special attention is devoted to the study of effects of the first-order perturbation formulas on the position of acoustical axes in weakly anisotropic media. Acoustical axes are a characteristic feature of anisotropic media. We can distinguish two kinds of acoustical axes (Schoenberg and Helbig, 1997). The first are the well-known *singular directions* in which the two qS waves propagating in anisotropic media propagate with the same phase velocity. These directions correspond to the places on the slowness surfaces at which the two qS-wave sheets are in a contact. Characteristic features of singular directions and their vicinity are distortions of slowness surfaces and strong variations of qS-wave polarization vectors (Schoenberg and Helbig, 1997; Shuvalov and Every, 1997; Vavryčuk, 2003). The second kind of acoustical axes are *longitudinal directions*. A characteristic feature of

[a]Electronic mail: farra@ipgp.jussieu.fr
[b]Author to whom correspondence should be addressed. Electronic mail: ip@ig.cas.cz
[c]Address used for correspondence.

longitudinal directions is propagation of purely longitudinal and purely transverse waves along them, i.e., propagation of qP waves with polarization parallel to the slowness vector (specifying the longitudinal direction) and propagation of qS waves with polarization perpendicular to the slowness vector. Helbig (1993) and Schoenberg and Helbig (1997) propose to use the acoustical axes for inversion of observed data into elastic parameters specifying anisotropy of the medium. It thus seems that importance of the acoustical axes might increase in the near future. In this contribution, we study effects of frequently used approximations on the accuracy with which the acoustical axes can be determined.

In Sec. II, basic equations are given. Properties of the matrix B_{mn} and its sensitivity to the parameters of a medium are studied. Alternative auxiliar vector bases, in which the matrix B_{mn} can be specified, are proposed and discussed. Then basic results of Farra (2001) are briefly reviewed and extended. It is shown that the higher-order ($n \geqslant 2$) perturbation formulas for qS waves are obtained separately for $qS1$ and $qS2$ waves. An alternative determination of two unit vectors specifying the qS-wave polarization, which leads to simplified polarization formulas, is proposed. In Sec. III, the conditions for the determination of acoustical axes in the first-order approximation of the perturbation theory are specified and discussed. The accuracy of the determination of acoustical axes determined by the perturbation formulas is studied. In Sec. IV, accuracy of perturbation formulas is illustrated on several numerical examples of media used in literature. The Appendix contains explicit expressions for the elements for one of many possible specifications of the matrix B_{mn}.

In the following, component notation is used. All the Roman lowercase indices range over the values 1, 2, and 3 and the uppercase indices over the values 1 and 2. Einstein summation convention is used for the repeated subscripts. Where necessary, superscripts are used. The superscripts in parentheses indicate an order of the quantity. In the matrices $\hat{B}^{(j)}_{mn}$ and in the vectors $\hat{e}^{(j)m}_i$, the upper indices j in parentheses indicate that $\hat{B}^{(j)}_{mn}$ and $\hat{e}^{(j)m}_i$ are used in expressions for the quantities of the jth or $(j+1)$-st order (see the text). The superscripts in the curly brackets (for example, $g^{\{m\}}_i$) indicate the type of the wave, $m=1$, 2 for the qS waves and $m=3$ for the qP wave. Voigt notation $A_{\alpha\beta}$ for density-normalized elastic parameters, with α, β running from 1 to 6, is used in parallel with the tensor notation a_{ijkl}.

II. HIGHER-ORDER PERTURBATION FORMULAS

We first study properties of the matrix B_{mn} playing the basic role in perturbation formulas for many wave attributes. Then we briefly review the formulas for the calculation of higher-order terms for the phase velocity and polarization of qP and qS waves derived by Farra (2001), and present their alternative form for qS waves.

A. Basic equations

Let us start from the Christoffel matrix

$$\Gamma_{jk} = a_{ijkl} n_i n_l . \tag{1}$$

The symbol a_{ijkl} denotes a tensor of density-normalized elastic parameters and n_i is a unit wave vector. The Christoffel matrix Γ_{jk} is symmetric and positive definite, see, e.g., Červený (2001). Thus for any direction of the wave vector n_i, Γ_{jk} has three positive and real eigenvalues $G_m = G_m(n_j)$ and three corresponding eigenvectors $g^{\{m\}}_i(n_j)$. They satisfy the following system of equations:

$$(\Gamma_{jk} - G_m \delta_{jk}) g^{\{m\}}_k = 0. \tag{2}$$

Two eigenvalues G_1 and G_2 with corresponding eigenvectors $g^{\{1\}}_i$ and $g^{\{2\}}_i$ correspond to the two qS waves, and the remaining eigenvalue G_3 with the eigenvector $g^{\{3\}}_i$ belongs to the qP wave. The eigenvalues are related to the squares of corresponding phase velocities $c_m(n_j)$,

$$G_m(n_j) = c^2_m(n_j). \tag{3}$$

The eigenvectors $g^{\{m\}}_i$ specify polarization vectors of the corresponding waves. The eigenvalues G_1 and G_2 are for most geological materials smaller than G_3 (Schoenberg and Helbig, 1997). Since the relation between G_1 and G_2 can be chosen arbitrarily, we choose it as follows: $G_2 \leqslant G_1 < G_3$.

The tensor a_{ijkl} can be expressed as follows:

$$a_{ijkl} = a^{(0)}_{ijkl} + \Delta a_{ijkl} . \tag{4}$$

Here $a^{(0)}_{ijkl}$ is a tensor of density-normalized elastic parameters in a reference isotropic medium:

$$a^{(0)}_{ijkl} = (\alpha^2 - 2\beta^2)\delta_{ij}\delta_{kl} + \beta^2(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \tag{5}$$

and Δa_{ijkl} is its perturbation. In (5), α and β denote P- and S-wave velocities of the reference isotropic medium.

B. Matrix B_{mn}

We introduce three mutually perpendicular unit vectors e^1_i, e^2_i, and e^3_i so that $e^3_i = n_i$. The vectors e^1_i and e^2_i, situated in the plane perpendicular to e^3_i, can be chosen arbitrarily.

We use the vectors e^k_i to define a matrix B_{mn},

$$B_{mn} = \Gamma_{jk} e^m_j e^n_k . \tag{6}$$

The matrix B_{mn} is independent of the choice of the reference isotropic medium. It is related to the *weak anisotropy* (WA) matrix $\mathcal{B}_{mn} = \Delta a_{ijkl} n_i n_l e^m_j e^n_k$, which, in contrast to B_{mn}, depends on the choice of the reference medium. Equation (13) of Pšenčík and Gajewski (1998) yields the relation between the two matrices:

$$B_{mn} = \mathcal{B}_{mn} + c^2_0 \delta_{mn} . \tag{7}$$

The symbol c_0 in Eq. (7) stands for the phase velocity of the reference isotropic medium. For $m=n=1$ or 2, $c_0 = \beta$; for $m=n=3$, $c_0 = \alpha$. From (7), we can see that the off-diagonal terms of the matrix B_{mn} are of the first order.

In the following we frequently use two specifications of the vectors e^K_i and, thus, of the matrix B_{mn}. In one we use a matrix $\hat{B}^{(0)}_{mn}$ defined by Eq. (6) with vectors e^1_i and e^2_i substituted by vectors $\hat{e}^{(0)1}_i$ and $\hat{e}^{(0)2}_i$ chosen so that

$$\hat{B}^{(0)}_{12} = 0, \quad \hat{B}^{(0)}_{11} > \hat{B}^{(0)}_{22} . \tag{8}$$

The second condition in Eq. (8) excludes singular directions from our considerations. The singular directions are studied in Sec. III. The diagonal elements of the matrix $\hat{B}^{(0)}_{mn}$ specify the first-order approximations $G^{(1)}_m$ of the eigenvalues of the Christoffel matrix, see Farra (2001). Specifically,

$$G^{(1)}_1=\hat{B}^{(0)}_{11}, \quad G^{(1)}_2=\hat{B}^{(0)}_{22}, \quad G^{(1)}_3=\hat{B}^{(0)}_{33}=B_{33}. \tag{9}$$

Since the elements of the matrix B_{mn} are independent of the velocities α and β of the reference medium, $G^{(1)}_m$ in (9) are independent of them, too.

In the other specification, we substitute the vectors e^K_i by vectors $\overline{e}^K_i$ given by Eqs. (A1) and (A2) of the Appendix. We denote the corresponding matrix B_{mn} by a bar: $\overline{B}_{mn}$. Explicit expressions for the elements of the matrix $\overline{B}_{mn}$ can be found in Eqs. (A4) in the Appendix. Elements of a matrix $\underline{B}_{mn}$ specified for arbitrarily chosen vectors e^J_i are related to B_{mn} by simple linear relations. If we denote by ϕ the acute angle between the vectors e^1_i and $\overline{e}^1_i$, the elements of the matrix $\overline{B}_{mn}$ transform into B_{mn} in the following simple way, see, e.g., Pšenčík and Vavryčuk (2002):

$$\begin{aligned}
B_{11}&=\overline{B}_{11}\cos^2\phi+2\overline{B}_{12}\cos\phi\sin\phi+\overline{B}_{22}\sin^2\phi,\\
B_{22}&=\overline{B}_{11}\sin^2\phi-2\overline{B}_{12}\cos\phi\sin\phi+\overline{B}_{22}\cos^2\phi,\\
B_{12}&=(\overline{B}_{22}-\overline{B}_{11})\cos\phi\sin\phi+\overline{B}_{12}(\cos^2\phi-\sin^2\phi),\\
B_{13}&=\overline{B}_{13}\cos\phi+\overline{B}_{23}\sin\phi,\\
B_{23}&=-\overline{B}_{13}\sin\phi+\overline{B}_{23}\cos\phi, \quad B_{33}=\overline{B}_{33}.
\end{aligned} \tag{10}$$

We can see that the elements $\overline{B}_{11}$, $\overline{B}_{12}$, and $\overline{B}_{22}$ transform into the elements of the matrix B_{mn} with the same subscripts. The same holds for the elements B_{13} and B_{23}. The element $\overline{B}_{33}$ is unaffected by the rotation.

In what follows it is shown that in the first-order approximation the elements of the above matrices control phase velocities, polarizations, and orientations of acoustical axes of elastic waves propagating in weakly anisotropic media. The elements B_{11}, B_{12}, and B_{22} are related to the qS-wave phase velocities and control orientation of singular directions. The element B_{33} is related to the qP-wave phase velocity. The elements B_{13} and B_{23} are related to the polarization vectors of qP and qS waves and control orientation of the longitudinal directions. Since, for given e^K_i, the elements of the matrix B_{mn} are linear functions of the WA or elastic parameters, it is desirable to deduce which parameters can affect individual elements of the matrix B_{mn}, and which parameters and in which form (individually or in a combination with others) can be retrieved from individual elements of the matrix B_{mn}.

By a simple inspection of Eqs. (A4) in the Appendix, we can find that the elements $\overline{B}_{11}$, $\overline{B}_{12}$, and $\overline{B}_{22}$ are controlled by 15 mutually independent coefficients, which depend linearly on all 21 WA parameters. The coefficients are $\epsilon_{15}-\epsilon_{35}$, $\epsilon_{24}-\epsilon_{34}$, $\chi_x-\epsilon_{34}$, $\chi_y-\epsilon_{35}$, $\epsilon_x-\delta_x+\epsilon_z$, $\epsilon_y-\delta_y+\epsilon_z$, $\delta_z-\epsilon_x-\epsilon_y$, $\chi_z-\epsilon_{16}$, $\chi_z-\epsilon_{26}$, γ_x, γ_y, γ_z, ϵ_{45}, ϵ_{46}, and ϵ_{56}. We can see that from the above elements of the matrix $\overline{B}_{mn}$ we can retrieve individually only six WA parameters. All remaining 15 WA parameters can be found only in combinations with others.

By a similar inspection of the element $\overline{B}_{33}$ we can find that the element is controlled by 15 different mutually independent coefficients, which depend on 15 qP-wave WA parameters. The coefficients are ϵ_x, ϵ_y, ϵ_z, $\epsilon_x-\delta_x+\epsilon_z$, $\epsilon_y-\delta_y+\epsilon_z$, $\delta_z-\epsilon_x-\epsilon_y$, χ_x, χ_y, χ_z, ϵ_{15}, ϵ_{16}, ϵ_{24}, ϵ_{26}, ϵ_{34}, and ϵ_{35}. We can see that in this case all 15 involved WA parameters can be determined from $\overline{B}_{33}$ individually.

The elements $\overline{B}_{13}$ and $\overline{B}_{23}$ are controlled by 14 mutually independent coefficients, which depend on 15 qP-wave WA parameters. The coefficients have the same form as for $\overline{B}_{33}$ except ϵ_x, ϵ_y, ϵ_z. Instead of these three coefficients we have their two combinations: $\epsilon_x-\epsilon_z$ and $\epsilon_y-\epsilon_z$. This means that from the elements $\overline{B}_{13}$ and $\overline{B}_{23}$ we can retrieve only nine individual WA parameters. The remaining six can be determined only in combinations. This has been observed by Zheng and Pšenčík (2002) during their inversion of synthetic polarization data in a VTI model.

A similar analysis in terms of the elastic parameters $A_{\alpha\beta}$ yields the following conclusions, which can be simply deduced from Eqs. (A3). The elements $\overline{B}_{11}$, $\overline{B}_{12}$, and $\overline{B}_{22}$ depend on all 21 elastic parameters. Only six of them, A_{44}, A_{55}, A_{66}, A_{46}, A_{56}, and A_{45}, can be found individually. The remaining 15 elastic parameters can be found only in combinations with other parameters.

The element $\overline{B}_{33}$ depends also on all 21 elastic parameters. But only nine of them, A_{11}, A_{22}, A_{33}, A_{15}, A_{16}, A_{24}, A_{26}, A_{34} and A_{35}, can be found individually. The remaining 12 elastic parameters can be found only in combinations with others.

The elements $\overline{B}_{13}$ and $\overline{B}_{23}$ depend again on all 21 elastic parameters. But only six of them, A_{15}, A_{16}, A_{24}, A_{26}, A_{34} and A_{35}, can be determined individually. The remaining 15 elastic parameters can be obtained only in combinations.

C. *qP* wave

We follow Farra (2001) and express the polarization vector $g^{\{3\}}_i$, related to the qP wave, in the vector basis $(\hat{e}^{(0)1}_i,\hat{e}^{(0)2}_i,e^3_i)$. Inserting $g^{\{3\}}_i$ specified in this way into (2) and multiplying the resulting equation successively by $\hat{e}^{(0)1}_i$, $\hat{e}^{(0)2}_i$, and e^3_i, we get a set of three equations from which we can determine a vector parallel to the exact polarization vector $g^{\{3\}}_i$:

$$\frac{\hat{B}^{(0)}_{13}}{G_3-G^{(1)}_1}\hat{e}^{(0)1}_i+\frac{\hat{B}^{(0)}_{23}}{G_3-G^{(1)}_2}\hat{e}^{(0)2}_i+e^3_i. \tag{11}$$

Note that the vector (11) is generally not unit. Equation (11) is derived under the assumption that the expressions in the denominators are nonzero. As mentioned above, in most geological materials this is guaranteed.

The vector (11) specifies the exact direction of polarization vector of the qP wave. Since the determination of the vectors $\hat{e}^{(0)K}_j$ and e^3_j of the elements of the matrix $\hat{B}^{(0)}_{mn}$ and of the first-order approximations of the eigenvalues is straightforward, only knowledge of the exact eigenvalue G_3 is nec-

essary for the determination of the direction of the polarization vector $g_i^{\{3\}}$ from Eq. (11). Equation (11) can be also used for the determination of the directions of the polarization vectors of different order of approximation, see Farra (2001). In the zeroth-order approximation, the polarization vector is e_i^3. In the first-order approximation we get

$$g_i^{(1)\{3\}}=\frac{\hat{B}_{13}^{(0)}}{G_3^{(0)}-G_1^{(0)}}\hat{e}_i^{(0)1}+\frac{\hat{B}_{23}^{(0)}}{G_3^{(0)}-G_2^{(0)}}\hat{e}_i^{(0)2}+e_i^3. \quad (12)$$

It is easy to show that in contrast to the general Eq. (11), Eq. (12) holds for arbitrarily chosen vectors e_i^K, not only for $\hat{e}_i^{(0)K}$. It can thus be rewritten in the form

$$g_i^{(1)\{3\}}=\frac{B_{13}}{\alpha^2-\beta^2}e_i^1+\frac{B_{23}}{\alpha^2-\beta^2}e_i^2+e_i^3. \quad (13)$$

We took into account that $G_3^{(0)}=\alpha^2$ and $G_1^{(0)}=G_2^{(0)}=\beta^2$. Equation (13) corresponds to the expression for the first-order perturbation of the qP-wave polarization vector of Jech and Pšenčík (1989). Since B_{13} and B_{23} are the first-order quantities, the vector $g_i^{(1)\{3\}}$ is a unit vector in the first-order approximation.

If we substitute $g_k^{\{m\}}$ in the Christoffel equation (2) by the vector (11), and multiply the resulting equation by e_j^3, we get

$$G_3=G_3^{(1)}+\frac{(\hat{B}_{13}^{(0)})^2}{G_3-G_1^{(1)}}+\frac{(\hat{B}_{23}^{(0)})^2}{G_3-G_2^{(1)}}. \quad (14)$$

This is a basic equation for the iterative determination of the higher-order approximations of the eigenvalue G_3, see Eq. (30) of Farra (2001). We can proceed in the same way as Farra (2001) to obtain approximations of G_3 of different orders. Specifically, in the first-order approximation we obtain

$$G_3^{(1)}=\hat{B}_{33}^{(0)}=B_{33}=\Gamma_{jk}e_j^3e_k^3, \quad (15)$$

see Eq. (9). It is interesting to note that an odd-order approximation of the eigenvalue G_3 can be determined from a lower odd-order approximation. The same holds for even-order approximations, see Farra (2001). The odd-order approximations of the eigenvalue G_3 are independent of the choice of the reference medium while the even-order approximations depend on its choice.

Since the elements B_{13}, B_{23}, and B_{33} appearing in Eqs. (13) and (15) are related only to the elements B_{13}, B_{23}, and B_{33}, see Eq. (10), from the discussion in Sec. II B it follows that the first-order approximations of the polarization vector (13) and of the eigenvalue (15) (representing square of the phase velocity of the qP wave in the first-order approximation) depend on 15 WA parameters. From the phase velocity all 15 WA parameters can be determined. Nine individual WA parameters and the remaining six in five combinations can be determined from the polarization vector. The polarization vector depends, in addition, on the difference of squared phase velocities in the reference isotropic medium: $\alpha^2-\beta^2$.

D. *qS* wave

We consider the two qS waves propagating in anisotropic media. We denote the faster wave $qS1$ and associate it with the index $K=1$. The slower $qS2$ wave is associated with the index $K=2$. For the qS waves we must use a procedure designed for the degenerate cases. In this section, we assume that the considered direction is not a singular direction. Singular directions are considered in Sec. III. We again follow Farra (2001) and express the polarization vector of the Kth qS wave, $g_i^{\{K\}}$, in the basis (e_i^1,e_i^2,e_i^3). Note that arbitrarily chosen vectors e_i^J are considered. Inserting $g_i^{\{K\}}$ specified in this way into (2) and multiplying the resulting equation successively by e_i^1, e_i^2, and e_i^3, we get, after some algebra, the following set of three equations for the projections of the vector $g_i^{\{K\}}$ into e_i^1, e_i^2, and e_i^3:

$$(M_{11}^{\{K\}}-G_K)g_m^{\{K\}}e_m^1+M_{12}^{\{K\}}g_m^{\{K\}}e_m^2=0, \quad (16a)$$

$$M_{12}^{\{K\}}g_m^{\{K\}}e_m^1+(M_{22}^{\{K\}}-G_K)g_m^{\{K\}}e_m^2=0, \quad (16b)$$

$$B_{13}g_m^{\{K\}}e_m^1+B_{23}g_m^{\{K\}}e_m^2+(G_3^{(1)}-G_K)g_m^{\{K\}}e_m^3=0. \quad (16c)$$

Here we use the notation of Farra (2001):

$$M_{ij}^{\{K\}}=B_{ij}+\frac{B_{i3}B_{j3}}{G_K-G_3^{(1)}}. \quad (17)$$

Note that $M_{ij}^{\{K\}}$ in Eq. (17) depends, through the index K, on the type of the considered qS wave. From the condition of solvability of Eqs. (16a) and (16b), we get expressions for the eigenvalues G_K of the two qS waves:

$$G_1=\tfrac{1}{2}[M_{11}^{\{1\}}+M_{22}^{\{1\}}+\sqrt{D_1}] \quad (18a)$$

and

$$G_2=\tfrac{1}{2}[M_{11}^{\{2\}}+M_{22}^{\{2\}}-\sqrt{D_2}], \quad (18b)$$

where

$$D_K=(M_{11}^{\{K\}}-M_{22}^{\{K\}})^2+4(M_{12}^{\{K\}})^2. \quad (19)$$

From Eqs. (16) and (17), we can determine vectors parallel to the exact polarization vectors $g_i^{\{1\}}$ and $g_i^{\{2\}}$. They can be written as follows:

$$P_Ke_i^1+Q_Ke_i^2+R_Ke_i^3. \quad (20)$$

The symbols P_K and Q_K in (20) have the following meaning:

$$P_1=\sqrt{\frac{1}{2}\left(1+\frac{M_{11}^{\{1\}}-M_{22}^{\{1\}}}{\sqrt{D_1}}\right)},$$
$$Q_1=\mathrm{sgn}(M_{12}^{\{1\}})\sqrt{\frac{1}{2}\left(1-\frac{M_{11}^{\{1\}}-M_{22}^{\{1\}}}{\sqrt{D_1}}\right)}, \quad (21a)$$

and

$$P_2=-\mathrm{sgn}(M_{12}^{\{2\}})\sqrt{\frac{1}{2}\left(1-\frac{M_{11}^{\{2\}}-M_{22}^{\{2\}}}{\sqrt{D_2}}\right)},$$
$$Q_2=\sqrt{\frac{1}{2}\left(1+\frac{M_{11}^{\{2\}}-M_{22}^{\{2\}}}{\sqrt{D_2}}\right)}. \quad (21b)$$

The symbol R_K in (20) reads

$$R_K=\frac{B_{13}P_K+B_{23}Q_K}{G_K-G_3^{(1)}}. \tag{21c}$$

As in the case of qP waves, the vectors (20) are generally not unit.

Equations (18) and (20), which hold for arbitrarily chosen vectors e_i^1 and e_i^2, are basic equations for the iterative determination of the higher-order approximations of the eigenvalues and eigenvectors of the two qS waves propagating in anisotropic media, see Farra (2001). It is only necessary to find higher-order approximations $M_{ij}^{(n)\{K\}}$ of the matrix $M_{ij}^{\{K\}}$ and insert them into (18) and (20). From (17), the first-order approximation $M_{ij}^{(1)\{K\}}$ is obviously $M_{ij}^{(1)\{K\}}=B_{ij}$, and it is thus the same for both qS waves. The nth-order approximation $M_{ij}^{(n)\{K\}}$ of the matrix $M_{ij}^{\{K\}}$, for $n\geqslant 2$, is defined by Eq. (17) with G_K substituted by $G_K^{(n-2)}$, see Farra (2001). Let us consider, for example, the $qS1$ wave. Inserting $M_{ij}^{(n)\{1\}}$ into (18a) yields nth-order approximation $G_1^{(n)}$ of the eigenvalue G_1. Inserting $M_{ij}^{(n)\{1\}}$ into (20), specified for $K=1$, yields $(n-1)$-st-order approximation $g_i^{(n-1)\{1\}}$ of the polarization vector $g_i^{\{1\}}$. For the $qS2$ wave, the procedure is analogous. As in the case of qP wave, the odd-order approximations of the eigenvalues (18) are independent of the choice of the reference medium while the even-order approximations depend on it.

Let us again consider the $qS1$ wave, for example, and let us introduce vectors $\hat{e}_i^{(n-1)1}$, $\hat{e}_i^{(n-1)2}$ chosen in an analogous way to the vectors $\hat{e}_i^{(0)1}$, $\hat{e}_i^{(0)2}$ in Eq. (8). Specifically, let us choose them so that they define the nth-order approximation $\hat{M}_{ij}^{(n)\{1\}}$ of the matrix $M_{ij}^{\{1\}}$, which satisfies

$$\hat{M}_{12}^{(n)\{1\}}=0, \quad \hat{M}_{11}^{(n)\{1\}}>\hat{M}_{22}^{(n)\{1\}}. \tag{22}$$

We can see that for $n=1$, Eq. (22) reduces to Eq. (8) since $\hat{M}_{ij}^{(1)\{K\}}=\hat{B}_{ij}^{(0)}$. This is a consequence of the equation $M_{ij}^{(1)\{K\}}=B_{ij}$, which holds for arbitrarily chosen vectors e_i^J. Using vectors $\hat{e}_i^{(n-1)1}$ and $\hat{e}_i^{(n-1)2}$ chosen in the above described way, we get from (18a) and (19) the expression for the nth approximation $G_1^{(n)}$ of the eigenvalue G_1:

$$G_1^{(n)}=\hat{M}_{11}^{(n)\{1\}}. \tag{23}$$

In the first-order approximation $(n=1)$, Eq. (23) yields the first equation of (9):

$$G_1^{(1)}=\hat{M}_{11}^{(1)\{1\}}=\hat{B}_{11}^{(0)}. \tag{24}$$

The direction of the vector $g_i^{(n)\{1\}}$, which represents the nth-order approximation $(n\geqslant 1)$ of the polarization vector $g_i^{\{1\}}$ of the considered $qS1$ wave is specified by the vector

$$\hat{e}_i^{(n)1}+\frac{\hat{B}_{13}^{(n)}}{G_1^{(n-1)}-G_3^{(1)}}e_i^3. \tag{25}$$

Note again that the vector (25) is not unit. Equation (25) follows from Eq. (20) specified for $K=1$. For $n=0$, Eq. (20) specified for $K=1$ yields the vector $\hat{e}_i^{(0)1}$ as the zeroth-order approximation of the vector $g_i^{\{1\}}$. The symbol $\hat{B}_{kl}^{(n)}$ denotes a matrix defined by Eq. (6), in which the vectors e_i^J are substituted by the vectors $\hat{e}_i^{(n)J}$. For $n=0$, the matrix $\hat{B}_{kl}^{(n)}$ reduces to $\hat{B}_{kl}^{(0)}$ satisfying (8). In the first-order approximation $(n=1)$, Eq. (25) yields the direction of the vector $g_i^{(1)\{1\}}$,

$$\hat{e}_i^{(1)1}+\frac{\hat{B}_{13}^{(1)}}{\beta^2-\alpha^2}e_i^3. \tag{26}$$

Note that when using the vectors $\hat{e}_i^{(n)J}$, the nth-order approximation of the polarization vector $g_i^{\{1\}}$ has zero projection into the vector $\hat{e}_i^{(n)2}$. Since the vector $\hat{e}_i^{(1)1}$ used in (26) differs from $\hat{e}_i^{(0)1}$, Eq. (26) differs from a similar equation for the first-order perturbation of the $qS1$-wave polarization vector given by Jech and Pšenčík (1989).

In an analogous way, we can derive corresponding approximations of the eigenvalue G_2 and of the direction of the polarization vector $g_i^{\{2\}}$ of the $qS2$ wave. In Eq. (22), we must substitute $\hat{M}_{ij}^{(n)\{1\}}$ by $\hat{M}_{ij}^{(n)\{2\}}$. This defines a new set of vectors $\hat{e}_i^{(n-1)1}$ and $\hat{e}_i^{(n-1)2}$, which are generally different from the set of vectors $\hat{e}_i^{(n-1)1}$ and $\hat{e}_i^{(n-1)2}$ related to the $qS1$ wave. Then, Eq. (23) becomes $G_2^{(n)}=\hat{M}_{22}^{(n)\{2\}}$ and in Eq. (25), $\hat{e}_i^{(n)1}$, $\hat{B}_{13}^{(n)}$, and $G_1^{(n-1)}$ must be substituted by $\hat{e}_i^{(n)2}$, $\hat{B}_{23}^{(n)}$, and $G_2^{(n-1)}$, respectively.

The vectors $\hat{e}_i^{(n)J}$ corresponding to one of the considered qS waves, let us say the $qS1$ wave, can be expressed in terms of arbitrarily chosen, mutually perpendicular unit vectors e_i^1 and e_i^2 in the following way:

$$\begin{aligned}\hat{e}_i^{(n)1}&=e_i^1\cos\phi^{(n)}+e_i^2\sin\phi^{(n)},\\ \hat{e}_i^{(n)2}&=-e_i^1\sin\phi^{(n)}+e_i^2\cos\phi^{(n)}.\end{aligned} \tag{27}$$

The angle $\phi^{(n)}$ can be determined from the equation

$$\tan 2\phi^{(n)}=\frac{2M_{12}^{(n+1)\{1\}}}{M_{11}^{(n+1)\{1\}}-M_{22}^{(n+1)\{1\}}}. \tag{28}$$

Equation (28) follows from the first condition in Eq. (22) if we take into account that the matrix $M_{ij}^{(n+1)\{1\}}$ satisfies the same relations (10) as the matrix B_{ij}. The second condition in (22) guarantees unique determination of the angle $\phi^{(n)}$ from Eq. (28). For $n=0$, Eq. (28) yields

$$\tan 2\phi^{(0)}=\frac{2B_{12}}{B_{11}-B_{22}}. \tag{29}$$

For the $qS2$ wave, we can proceed in the same way. We obtain, of course, generally different vectors $\hat{e}_i^{(n)1}$ and $\hat{e}_i^{(n)2}$ and different values of the angles $\phi^{(n)}$. Only in the zeroth-order [see Eq. (29)] and the first-order approximation, $\phi^{(0)}$ and $\phi^{(1)}$ have universal meaning for both qS waves since $M_{ij}^{(1)\{1\}}=M_{ij}^{(1)\{2\}}$ and $M_{ij}^{(2)\{1\}}=M_{ij}^{(2)\{2\}}$.

Let us mention that the procedure of selection of vectors $\hat{e}_i^{(n-1)J}$ satisfying Eq. (22) can be extended to the limiting case. For the Kth qS wave we can specify vectors $\hat{e}_i^J$ so that the elements of the matrix $\hat{M}_{ij}^{\{K\}}$ [the matrix $\hat{M}_{ij}^{\{K\}}$ is the exact matrix $M_{ij}^{\{K\}}$, see (17), specified for the vectors $\hat{e}_i^J$] satisfy similar relations to those in Eq. (22):

$$\hat{M}_{12}^{\{K\}}=0, \quad \hat{M}_{11}^{\{K\}}>\hat{M}_{22}^{\{K\}}. \tag{30}$$

For such a specification, we get for the $qS1$ wave from (18a) and (19) specified for $K=1$,

$$G_1=\hat{M}_{11}^{\{1\}}, \tag{31}$$

and for the direction of the corresponding exact polarization vector,

$$\hat{e}_i^1+\frac{\hat{B}_{13}}{G_1-G_3^{(1)}}e_i^3. \tag{32}$$

The matrix $\hat{B}_{mn}$ is specified by the vectors $\hat{e}_i^J$. Again, analogous expressions for the eigenvalue G_2 and for the direction of the polarization vector $g_i^{\{2\}}$ can be written for the $qS2$ wave. Let us again emphasize that the corresponding vectors $\hat{e}_i^1$ and $\hat{e}_i^2$ are generally different from the vectors $\hat{e}_i^J$ related to the $qS1$ wave.

The first-order approximation of the qS-wave polarization vectors [see (26) for $qS1$ wave, for example] depends on the elements B_{13} and B_{23} of the matrix B_{mn} [the elements $\hat{B}_{13}^{(1)}$ and $\hat{B}_{23}^{(1)}$ are functions of B_{13} and B_{23} only, see Eq. (10)] and, through the angle $\phi^{(1)}$, which controls specification of the vectors $\hat{e}_i^{(1)J}$ [see (28) specified for $n=1$], also on elements B_{11}, B_{12}, and B_{22}. The first-order approximation of the eigenvalues (24) depends on elements B_{11}, B_{12}, and B_{22} [since $\hat{B}_{11}^{(0)}$ and $\hat{B}_{22}^{(0)}$ depend on B_{11}, B_{12}, and B_{22} only, see Eq. (10)]. From the discussion in Sec. II B, it follows that the first-order approximations of the eigenvalues (representing squares of the phase velocities in the first-order approximations) and of the polarization vectors of qS waves depend on the complete set of 21 WA parameters. Six individual WA parameters and remaining 15 in nine combinations can be determined from the phase velocity. Fifteen individual WA parameters and remaining six in five combinations can be determined from the polarization vectors. The polarization vectors are, in addition, also controlled by the difference of squared velocities of the reference isotropic medium: $\alpha^2-\beta^2$.

III. ACOUSTICAL AXES

Characteristic feature of any anisotropic medium are so-called *acoustical axes*. According to Schoenberg and Helbig (1997) the acoustical axes specify two kinds of directions: longitudinal and singular.

A. Longitudinal directions

A longitudinal direction is a direction in which the qP wave is purely longitudinal and the qS waves are purely transverse. The polarization vector $g_i^{\{3\}}$ of the qP wave is parallel to the corresponding wave normal n_i; the polarization vectors $g_i^{\{K\}}$ of the qS waves are perpendicular to n_i. By taking this into account in Eqs. (11) and in the expression for R_K in (21c), we can see that in the longitudinal direction

$$\hat{B}_{13}^{(0)}=\hat{B}_{23}^{(0)}=B_{13}=B_{23}=0. \tag{33}$$

We can see that Eq. (33) holds for arbitrary choice of vectors e_i^K. From Eqs. (12) or (13), we can see that if Eq. (33) holds, also the first-order approximation of the qP-wave polarization vector is parallel to the wave vector n_i. From Eq. (26), we can see that the first-order approximation of the $qS1$-wave polarization vector is perpendicular to the wave vector n_i if Eq. (33) holds. Thus the longitudinal directions determined from the first-order perturbation formulas are identical with exact longitudinal directions.

Let us mention that the condition (33) is automatically satisfied in isotropic media. This means that every direction in an isotropic medium is a longitudinal direction.

From Eq. (14) we can derive another interesting property of the first-order perturbation of the qP-wave eigenvalue. For $G_3>G_1^{(1)}$ and $G_3>G_2^{(1)}$, Eq. (14) indicates that $G_3\geqslant G_3^{(1)}$, i.e., the qP-wave first-order formula yields a value which is less than or at most equal to the exact eigenvalue. An alternative proof of this inequality can be found in the Appendix of Pšenčík and Gajewski (1998). We can see from Eq. (14) that the equality $G_3=G_3^{(1)}$ occurs when (33) holds, i.e., in longitudinal directions.

A similar property as above holds also for the qS waves. The proof is only slightly more complicated. Let us consider vectors $\hat{e}_i^{(0)J}$ chosen so that conditions (8) are satisfied and let us rewrite Eqs. (18) for this specification. We get

$$\begin{aligned} G_1&=\frac{1}{2}\left(\hat{B}_{11}^{(0)}+\hat{B}_{22}^{(0)}+\frac{(\hat{B}_{13}^{(0)})^2+(\hat{B}_{23}^{(0)})^2}{G_1-G_3^{(1)}}+\sqrt{D_1}\right),\\ G_2&=\frac{1}{2}\left(\hat{B}_{11}^{(0)}+\hat{B}_{22}^{(0)}+\frac{(\hat{B}_{13}^{(0)})^2+(\hat{B}_{23}^{(0)})^2}{G_2-G_3^{(1)}}-\sqrt{D_2}\right). \end{aligned} \tag{34}$$

Let us also rewrite Eq. (19) for the same specification:

$$\begin{aligned} D_K=&(\hat{B}_{11}^{(0)}-\hat{B}_{22}^{(0)})^2+\left(\frac{(\hat{B}_{13}^{(0)})^2+(\hat{B}_{23}^{(0)})^2}{G_K-G_3^{(1)}}\right)^2\\ &+2(\hat{B}_{11}^{(0)}-\hat{B}_{22}^{(0)})\frac{(\hat{B}_{13}^{(0)})^2-(\hat{B}_{23}^{(0)})^2}{G_K-G_3^{(1)}}. \end{aligned} \tag{35}$$

The quantity D_K satisfies the following inequalities:

$$\left(\hat{B}_{11}^{(0)}-\hat{B}_{22}^{(0)}+\frac{(\hat{B}_{13}^{(0)})^2+(\hat{B}_{23}^{(0)})^2}{G_K-G_3^{(1)}}\right)^2\leqslant D_K, \tag{36a}$$

$$D_K\leqslant\left(\hat{B}_{11}^{(0)}-\hat{B}_{22}^{(0)}-\frac{(\hat{B}_{13}^{(0)})^2+(\hat{B}_{23}^{(0)})^2}{G_K-G_3^{(1)}}\right)^2. \tag{36b}$$

The inequalities in (36) follow from

$$\begin{aligned} -[(\hat{B}_{13}^{(0)})^2+(\hat{B}_{23}^{(0)})^2]&\leqslant(\hat{B}_{13}^{(0)})^2-(\hat{B}_{23}^{(0)})^2\\ &\leqslant(\hat{B}_{13}^{(0)})^2+(\hat{B}_{23}^{(0)})^2 \end{aligned} \tag{37}$$

and from $G_K<G_3^{(1)}$. Inserting (36b) to the expression for G_1 in (34) and (36a) to G_2, we get two important results:

$$G_1\leqslant\hat{B}_{11}^{(0)}=G_1^{(1)},\quad G_2\leqslant\hat{B}_{22}^{(0)}=G_2^{(1)}. \tag{38}$$

We can see that the first-order formulas yield values which are greater than or at least equal to exact eigenvalues of the qS waves. As in the case of qP wave, we can see from Eqs. (17)–(19) that the equality $G_1=G_1^{(1)}$ and $G_2=G_2^{(1)}$ occurs when (33) holds, i.e., in longitudinal directions.

Let us mention that the equality of the first-order approximation of only one of the qS-wave eigenvalues to its exact counterpart does not imply a longitudinal direction. The SH wave in a TI medium can be taken as an example.

Equations (33) can be thus used for the determination of longitudinal directions if the WA parameters are known or, vice versa, for the determination (or constraint) of the WA parameters if the longitudinal directions are known from observations. In the (x,z) plane of an orthorhombic medium, the equation $\overline{B}_{13}=0$ yields two ''bound'' directions (Helbig, 1993) along coordinate axes x and z and an equation

$$2\epsilon_x n_1^2 - 2\epsilon_z n_3^2 + \delta_x (n_3^2 - n_1^2) = 0. \tag{39}$$

Equation $\overline{B}_{23}=0$ is satisfied automatically. For $\tan^2\theta = n_1^2/n_3^2$, where θ is measured from the z-axis, and under one of the conditions

$$\tfrac{1}{2}\delta_x < \min(\epsilon_x, \epsilon_z), \quad \tfrac{1}{2}\delta_x > \max(\epsilon_x, \epsilon_z), \tag{40}$$

Eq. (39) yields

$$\tan^2\theta = \frac{2\epsilon_z - \delta_x}{2\epsilon_x - \delta_x}. \tag{41}$$

Equation (41) and conditions (40) are equivalent to those given for the same situation by Schoenberg and Helbig (1997). In a similar way we get for the (x,y) plane of an orthorhombic medium from Eqs. (33) two ''bound'' directions along coordinate axes x and y and the direction

$$\tan^2\theta = \frac{2\epsilon_y - \delta_z}{2\epsilon_x - \delta_z}. \tag{42}$$

Here θ is measured from the y axis. The longitudinal directions can exist if one of the conditions

$$\tfrac{1}{2}\delta_z < \min(\epsilon_x, \epsilon_y), \quad \tfrac{1}{2}\delta_z > \max(\epsilon_x, \epsilon_y) \tag{43}$$

is satisfied.

B. Singular directions

The second type of acoustical axes is connected with singular directions. Singular directions exist only for qS waves and occur when the phase velocities (or the eigenvalues G_M) of the two qS waves are equal. This is the so-called degenerate situation in perturbation analysis. In such a situation, it is not possible to specify uniquely the polarization vectors of the qS waves. It is only possible to find the plane in which the polarization vectors are situated, i.e., the plane perpendicular to the third polarization vector. This fact reflects the indeterminacy of corresponding formulas for the polarization vectors. The expression (17) is independent of the type of the considered wave in the singular direction because $G_1=G_2$. Inspection of Eqs. (18) reveals that the equality $G_1=G_2$ occurs if $D_1=D_2=0$. Because D_K is given, see (19), by a sum of two quadratic terms, each of these terms must be zero. This yields

$$M_{11}^{\{K\}} = M_{22}^{\{K\}}, \quad M_{12}^{\{K\}} = 0. \tag{44}$$

Equation (44) holds simultaneously for $K=1$ and $K=2$, and for an arbitrary choice of the vectors e_i^K. In the first-order approximation, the singular direction is specified by

$$B_{11} = B_{22}, \quad B_{12} = 0. \tag{45}$$

In the (x,z) plane of an orthorhombic medium, the equation $B_{12}=0$ is automatically satisfied and equation $B_{11}=B_{22}$ has an explicit form

$$\begin{aligned}&\alpha^2(\epsilon_x - \delta_x + \epsilon_z)\cos^4\theta \\ &\quad + [\beta^2(\gamma_y - \gamma_z) - \alpha^2(\epsilon_x - \delta_x + \epsilon_z)]\cos^2\theta \\ &\quad + \beta^2(\gamma_z - \gamma_x) = 0.\end{aligned} \tag{46}$$

In the plane (y,z), the equation $B_{11}=B_{22}$ has the form

$$\begin{aligned}&\alpha^2(\epsilon_y - \delta_y + \epsilon_z)\cos^4\theta \\ &\quad + [\beta^2(\gamma_x - \gamma_z) - \alpha^2(\epsilon_y - \delta_y + \epsilon_z)]\cos^2\theta \\ &\quad + \beta^2(\gamma_z - \gamma_y) = 0.\end{aligned} \tag{47}$$

If a singular direction differs from a longitudinal direction, the equality $G_1=G_2$ does not imply generally $G_1^{(1)}=G_2^{(1)}$. This means that the singular direction in the first-order approximation, specified by (45), differs generally from the actual singular direction specified by (44). This also holds for approximations of higher orders. In the limit, the singular direction should converge to the actual one. If Eqs. (44) and (45) are satisfied for the same direction, this automatically implies that that direction is longitudinal. Let us mention that the conditions (44) and (45) are automatically satisfied in isotropic media. This means that every direction in an isotropic medium is a singular direction.

IV. NUMERICAL EXAMPLES

In applications, the most frequently used formulas are formulas based on the first-order approximation. In the following, we shall compare results of the first- and second-order perturbation formulas for qS waves with exact results for several types of anisotropic media. This will enable us to estimate accuracy of the first-order formulas in a vicinity of singular directions. Let us emphasize that the results of this section were obtained by using Eqs. (18) and (20). In order to calculate the nth-order approximation of the square of the phase velocity from Eq. (18), the nth-order approximation of the matrix $M_{ij}^{\{K\}}$ was used. The velocity was calculated as a square root from the corresponding eigenvalue. For the determination of the nth-order approximation of polarization vectors from (20), the $(n+1)$-th-order approximation of $M_{ij}^{\{K\}}$ was used.

For illustration of effects of a singularity on the first-order approximation formulas, let us consider a simple transversely isotropic medium with vertical axis of symmetry. We use the model of thin water-filled cracks of Shearer and Chapman (1989). The model is specified by the density-normalized elastic parameters A_{ij}, in $(\mathrm{km/s})^2$, with values $A_{11}=A_{22}=20.16$, $A_{33}=19.63$, $A_{12}=7.40$, $A_{13}=A_{23}=7.26$, $A_{44}=A_{55}=3.48$, and $A_{66}=6.38$. Anisotropy of this model is rather strong, about 29% and 30% for qS waves. Due to the axial symmetry of the medium, it is sufficient to investigate just a quadrant of a vertical plane containing the axis of symmetry. Figure 1 shows variations of phase velocities of the two qS waves in such a section. The angle θ specifies the

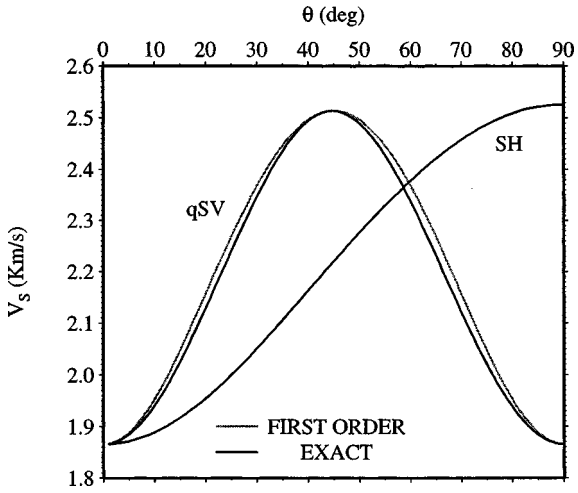


FIG. 1. *qSV*- and *SH*-wave phase velocity sections for the VTI medium specified in the text. Comparison of exact and first-order values.

direction of the wave vector, $\theta=0°$ corresponds to the direction along the axis of symmetry, and $\theta=90°$ corresponds to the direction perpendicular to it. The two distinct curves correspond to the *SH* and *qSV* waves. Polarization of the *SH* wave is perpendicular to the plane of propagation and thus to the wave vector (therefore we omit the letter *q* in front of *SH*); the *qSV* wave is polarized in the plane of propagation, its polarization is generally not perpendicular to the wave vector. The first-order phase velocity formula for the *SH* wave yields exact values of the phase velocity (because the zeroth-order approximation of the *SH*-wave polarization vector coincides with its exact counterpart). For the *qSV* wave the values generally differ, the approximate values being always greater than or equal to the exact ones (see the analytical proof above). Results of the approximate formula coincide with exact for $\theta=0°$ (kiss singularity), for approximately 45° (longitudinal direction) and for 90° (another longitudinal direction). For the goals of this paper, the most interesting is the intersection of the *SH*- and *qSV*-wave phase velocities at 58.59°, which corresponds to an intersection singularity. We can see that the singularity estimated by the first-order formula is shifted from the direction of the true singularity to the direction $\theta=59.7°$ determined from (46). The first-order perturbation formula thus yields distorted result. It is of interest that there is no shift in case of the kiss singularity. The kiss singularity direction coincides with the longitudinal direction, which is determined exactly by the first-order formulas.

Let us now consider a more complicated situation, an orthorhombic model used by Farra (2001). The model is specified by the density-normalized elastic parameters A_{ij}, in $(\text{km/s})^2$, with values $A_{11}=10.8$, $A_{22}=11.3$, $A_{33}=8.5$, $A_{12}=2.2$, $A_{13}=1.9$, $A_{23}=1.7$, $A_{44}=3.6$, $A_{55}=3.9$, and $A_{66}=4.3$. Squares of the *P*- and *S*-wave velocities of the reference isotropic medium are chosen 10.04 and 4.01 $(\text{km/s})^2$, respectively. Figure 2 shows exact (top), first-order (middle), and second-order (bottom) maps of relative differences (in %) of $qS1$ and $qS2$ phase velocities as functions of the polar angle θ and the azimuth φ specifying the wave vector.

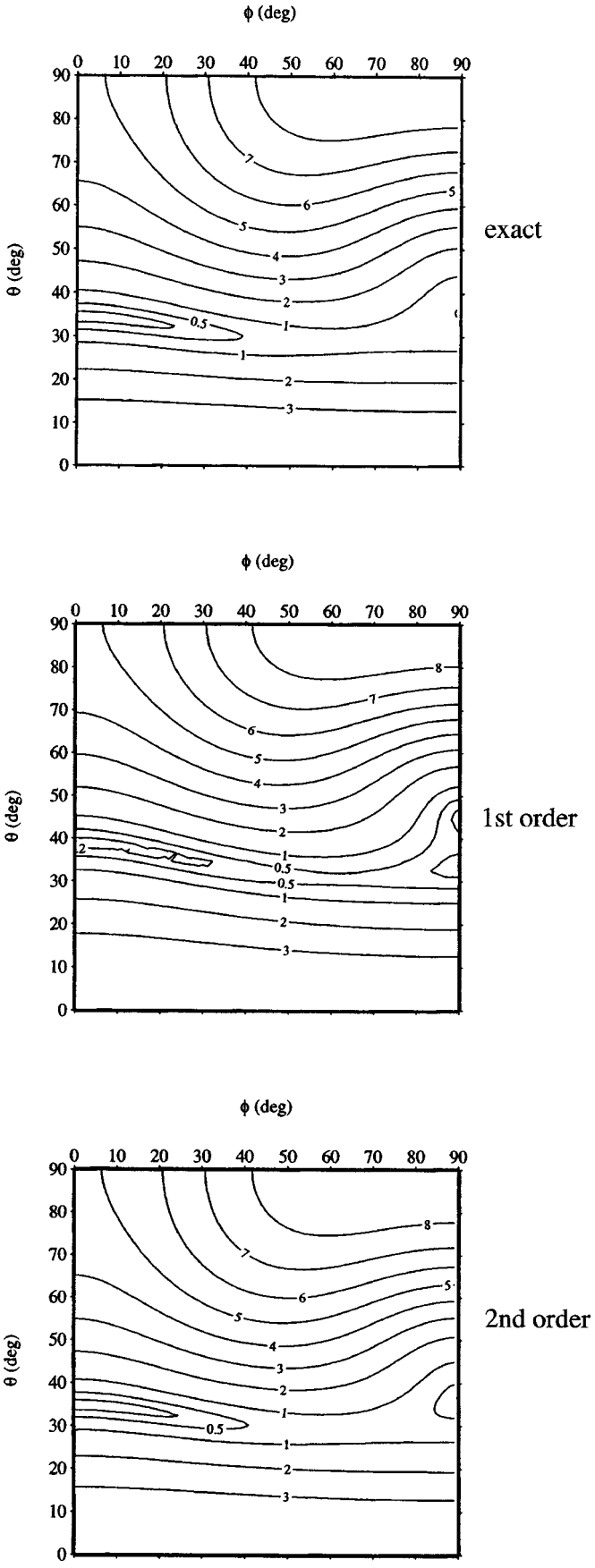


FIG. 2. Exact, first-order, and second-order maps of relative differences of $qS1$- and $qS2$-wave phase velocities as functions of azimuth ϕ and polar angle θ for an orthorhombic medium of Farra (2001).

The relative differences are calculated as differences of the qS-wave velocities normalized by their average. There is a point singularity at $\varphi=0°$ and $\theta=34.4°$ and the region of small differences extends to larger values of φ in the map of exact differences. Another region of small differences can be observed on the right-hand side of the exact map for $\varphi=90°$ and $\theta\sim35°$. Let us now compare the exact map with its first-order approximation in the middle frame. We can see several interesting phenomena. The most significant is the shift of the region of minimum differences to higher values of θ, from 34.4° to 38.91° following from (46). Another important feature is the appearence of two minima seemingly indicating existence of singularities for $\varphi=90°$, $\theta=33°$, and $\theta=45°$. The bottom of the valley crossing the plots for $\theta\sim30°-40°$ is deeper in the first-order map. For relative differences of $qS1$ and $qS2$ phase velocities greater than approximately 3%, the differences between the top and middle frames are minimum. From inspection of the frame of the second-order approximation, we can see that it gives effectively the same results as the exact formula. Slight exception is the region of $\varphi=90°$ and $\theta\sim35°$.

In the following, we show effects of the first-order perturbations on the model of orthorhombic medium proposed and studied in detail by Schoenberg and Helbig (1997). The model has been obtained by combining finely layered (VTI) model with vertical fractures. The model is specified by the density-normalized elastic parameters A_{ij}, in $(\mathrm{km/s})^2$, with values $A_{11}=9.00$, $A_{22}=9.84$, $A_{33}=5.94$, $A_{12}=3.60$, $A_{13}=2.25$, $A_{23}=2.40$, $A_{44}=2.00$, $A_{55}=1.60$, and $A_{66}=2.18$. Squares of the P- and S-wave velocities of the reference isotropic medium are chosen as 7.60 and 2.26 $(\mathrm{km/s})^2$, respectively. Figure 3 shows again exact (top), first-order (middle), and second-order (bottom) maps of relative differences of $qS1$ and $qS2$ phase velocities. This model has four singularities in the shown map: two singularities (conical points) for $\varphi=0°$ and $\theta=20.1°$ and $\theta=59.8°$, one conical point for $\varphi=90°$ and $\theta=72.5°$, and another one for $\varphi=44.89°$ and $\theta=46.53°$ (see Schoenberg and Helbig, 1997). Comparison of the exact and first-order maps yields several interesting phenomena. We can see that, as in Fig. 2, the singularity $\varphi=0°$, $\theta=59.8°$ is shifted within the plane of symmetry $\varphi=0°$ to $\theta=66.23°$ following from (46). The other singularity in the same plane, $\varphi=0°$, $\theta=20.1°$, remains, however, at nearly the same position, specifically at $\theta=20.09°$. This is a result of the closeness of the longitudinal direction specified by $\varphi=0°$ and $\theta=20.3°$. As mentioned above, the first-order formulas yield exact results for the longitudinal directions. Thus the differences between the first-order approximations of phase velocities of the qS waves in a vicinity of a longitudinal direction are also close to exact ones. The singularity $\varphi=90°$, $\theta=72.5°$ is shifted in the plane $\varphi=90°$ to $\theta=76.74°$ following from (47). Interesting is the shift of the point singularity $\varphi=44.89°$, $\theta=46.53°$ situated off the symmetry planes. The shift of the singularity in the first-order approximation is in both directions. It is about 2° in φ and about 4° in the polar angle θ. As in Fig. 2, the results of the second-order formula practically coincide with exact results. The only exception is the singularity $\varphi=90°$, $\theta=72.5°$, which remains shifted by about 2°. The use of the

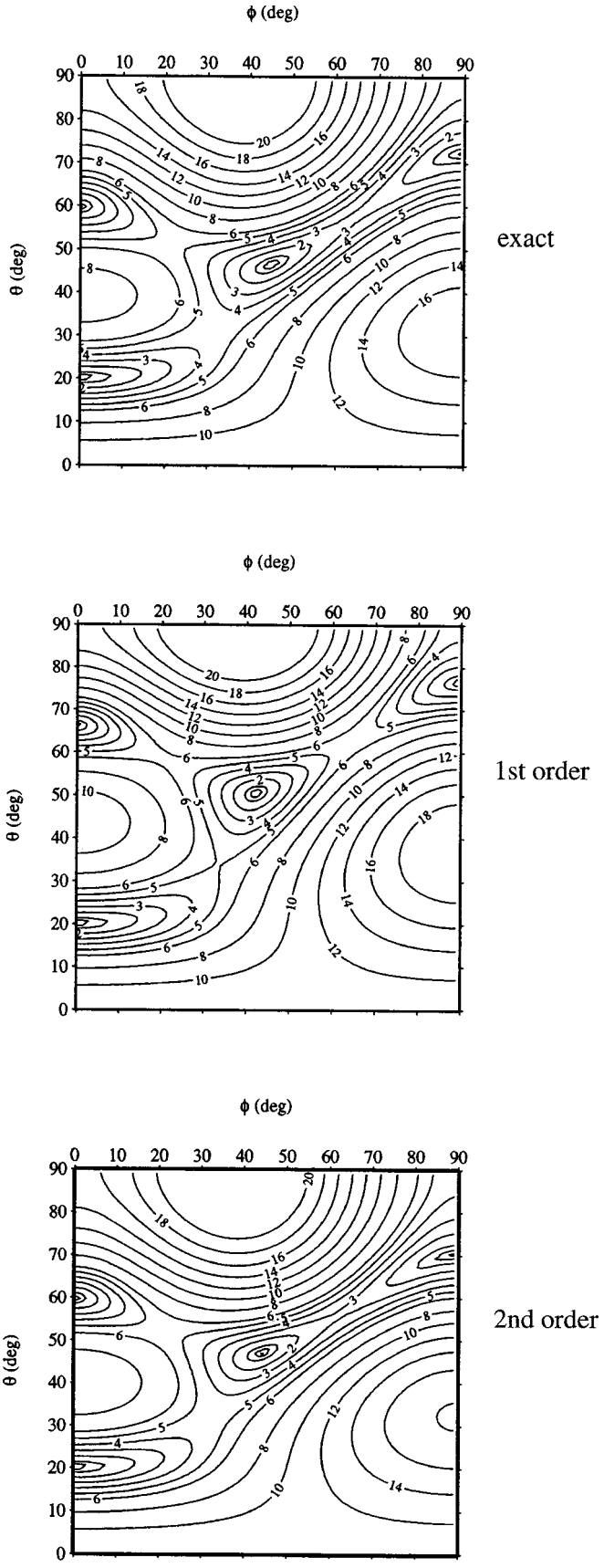


FIG. 3. Exact, first-order, and second-order maps of relative differences of $qS1$- and $qS2$-wave phase velocities as functions of azimuth ϕ and polar angle θ for an orthorhombic medium of Schoenberg and Helbig (1997).

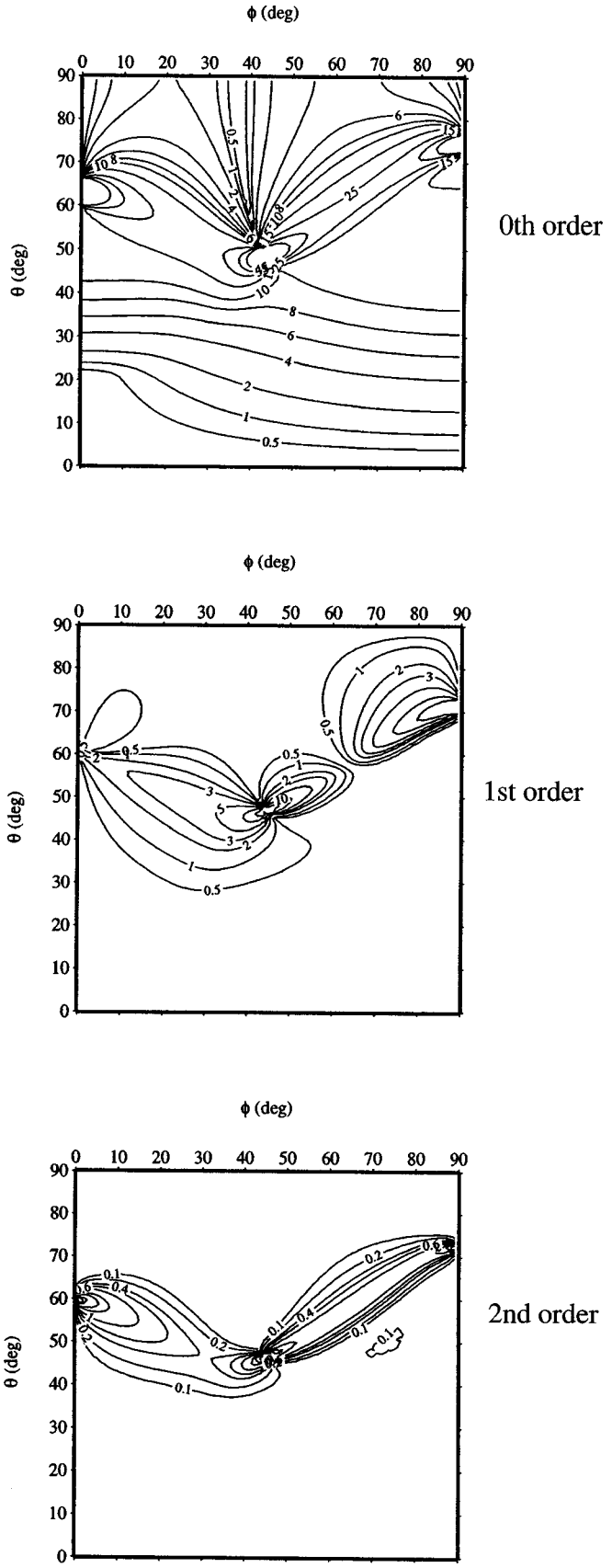


FIG. 4. Deviations (in degrees) of the exact $qS1$-wave polarization vector from its zeroth-, first- and second-order approximations as functions of azimuth ϕ and polar angle θ for an orthorhombic medium of Schoenberg and Helbig (1997).

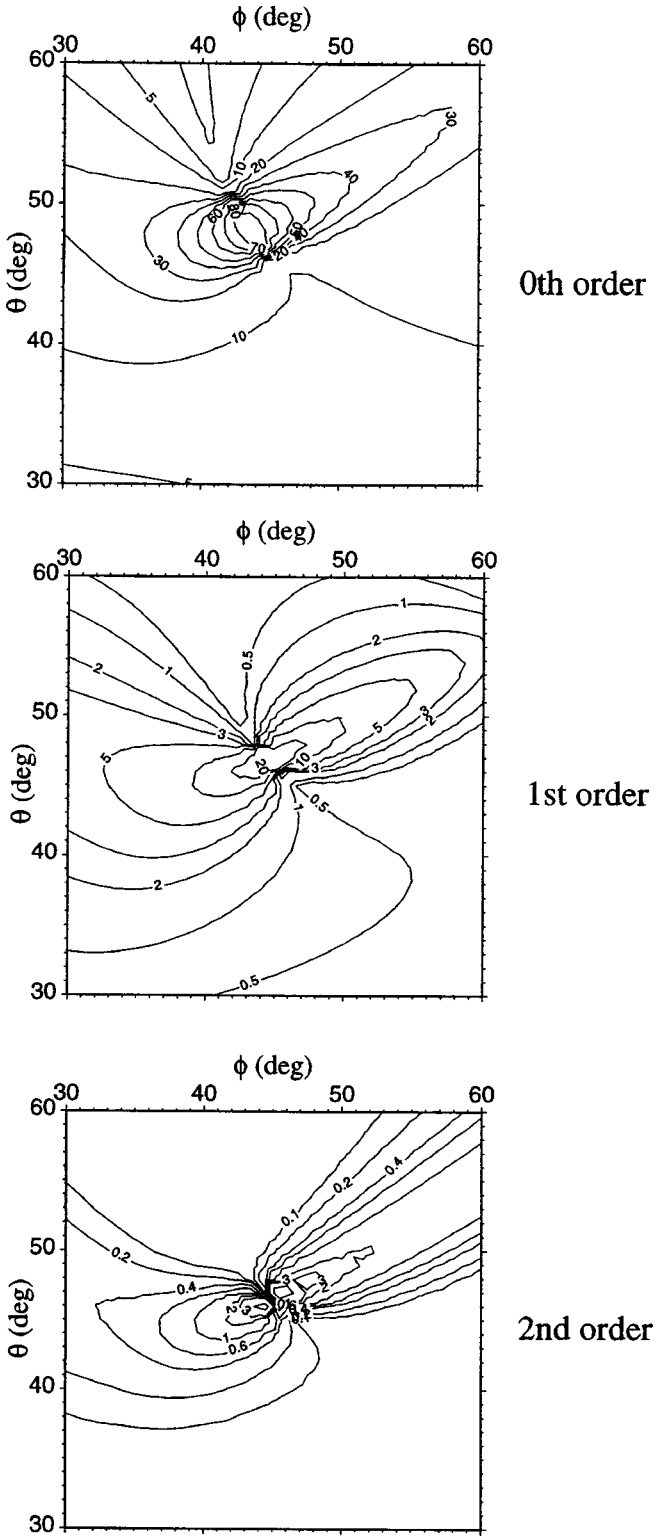


FIG. 5. The same as in Fig. 4 but for the azimuths and polar angles in the interval 30°–60°.

third-order formula would shift the singularity to the proper position.

The effect of singularities on polarization vectors is much more pronounced. In singular directions the formulas for polarization vectors are indetermined. In their vicinity, polarization vectors change fast. This can be seen in Fig. 4. Figure 4 shows deviations of the exact polarization vector of

the faster $qS1$ wave from the zeroth-order (top), first-order (middle) and second-order (bottom) approximation of the same polarization vector. From the upper plot we can see that the zeroth-order approximation yields deviation less than 1° in only small regions. Generally, the deviation of the zeroth-order approximation is large; it reaches values of 90°! The points in which the isolines concentrate represent actual and shifted positions of singularities. The actual positions are related to the exact polarization vector, which varies significantly in their vicinity. The shifted positions are related to the zeroth-order approximation of the polarization vector, which varies significantly in the vicinity of the shifted positions (see the top plot in Fig. 3). There is only very slightly increased variation of deviations in the vicinity of the singularity $\varphi=0°$, $\theta=20.1°$. The first-order approximation (middle plot) of the polarization vector reduces the deviations significantly. The maximum deviations are again up to 90°, but they are concentrated to the narrow region connecting three singularities. Except for the singularity $\varphi=90°$, $\theta=72.5°$, the concentrations of isolines are close to actual positions of singularities. This picture further improves when second-order approximation of the polarization vector is used (see the bottom plot in Fig. 4). The maximum deviation does not exceed 1° in most of the studied region; all non-negligible deviations are concentrated in the narrow belt connecting the three actual singularities whose positions are indicated by concentration of isolines.

Figure 5 shows detailed maps of the deviations of the exact polarization vector of the faster $qS1$ wave from the zeroth-order (top), first-order (middle) and second-order (bottom) approximation of the same polarization vector. We concentrate there on the vicinity of the singularity which can be seen in Fig. 4 in the region of φ and θ between 40° and 50°. The shift of the singularity with varying order of the approximation is clearly visible.

V. CONCLUSIONS

A detailed analysis of higher-order formulas of Farra (2001) for the eigenvalues and directions of the polarization vectors of qP and qS waves in weakly anisotropic media was performed. Possible ways of constructing separate higher-order perturbation for $qS1$ and $qS2$ waves were shown. The formulas depend on various elements of the matrix B_{mn}, which, in turn, depend linearly on the WA or elastic parameters of a medium. Thus the sensitivity analysis of attributes of elastic waves to the parameters of the medium reduces to the sensitivity analysis of the elements of the matrix B_{mn}.

Analysis of the first-order perturbation formulas for phase velocities and polarization vectors yielded the following results. From inversion of qP-wave phase velocity, 15 independent WA parameters can be found. If information about qP-wave polarization and/or orientation of longitudinal axes is available in addition, the same 15 WA parameters can be found, no more. From inversion of qS-wave velocities, only six independent WA parameters can be found. Addition of information about singular directions cannot change this situation. If, however, information about qS-wave polarization and/or of orientation of the longitudinal axes is available, the number of determinable independent WA parameters increases to 15. In order to determine all 21 WA parameters, observations of both qP- and qS-wave phase velocities are necessary. From observations of qP- and qS-wave polarizations alone, only nine individual WA parameters can be found. Additional use of the acoustical axes, specifically of the singular directions, adds six further independent WA parameters. Using (A3), a similar analysis can be made for elastic parameters.

Another goal of this contribution was to show the effect of the first-order perturbation formulas on the determination of acoustical axes. For this purpose, explicit analytical conditions for the determination of the acoustical axis were given. We have found, both analytically and in numerical examples, that the approximations do not affect positions of the longitudinal directions. They may, however, affect quite significantly positions of singularities. The effects can be removed by using higher-order approximations.

ACKNOWLEDGMENTS

This work was supported by the CNRS-AS ČR project, by consortium project ''Seismic waves in complex 3-D structures (SW3D)'' and by the Grant Agency of the Czech Republic under Contract No. 205/00/1350. Selected results were presented at the 10th International Workshop on Seismic Anisotropy (IWSA) in Tutzing, Germany, 14–19 April 2002. The authors are grateful to Peter Bakker and Lin-Ping Song for their comments.

APPENDIX: EXPRESSIONS FOR ELEMENTS OF THE MATRIX $\underline{B}_{mn}$

We introduce three mutually perpendicular unit vectors $\overrightarrow{e_i^1}$, $\overrightarrow{e_i^2}$, and $\overrightarrow{e_i^3}$ so that $\overrightarrow{e_i^3}=n_i$, where n_i is a unit wave vector. The vectors $\overrightarrow{e_i^1}$ and $\overrightarrow{e_i^2}$ are situated in the plane perpendicular to $\overrightarrow{e_i^3}$ and they are selected so that

$$\begin{aligned}&\overrightarrow{e_i^1}\equiv D^{-1}(n_1n_3,n_2n_3,n_3^2-1),\\&\overrightarrow{e_i^2}\equiv D^{-1}(-n_2,n_1,0),\quad \overrightarrow{e_i^3}=n_i\equiv(n_1,n_2,n_3),\end{aligned}\tag{A1}$$

where

$$D=(n_1^2+n_2^2)^{1/2},\quad n_1^2+n_2^2+n_3^2=1.\tag{A2}$$

In the following, a reference isotropic medium with the P-wave velocity α and S-wave velocity β is considered. We specify the symmetric matrix B_{mn} given in Eq. (6) for vectors (A1) and denote it $\underline{B}_{mn}$. The matrix B_{mn} can be expressed in terms of 21 WA parameters:

$$\epsilon_x=\frac{A_{11}-\alpha^2}{2\alpha^2},\quad \epsilon_y=\frac{A_{22}-\alpha^2}{2\alpha^2},\quad \epsilon_z=\frac{A_{33}-\alpha^2}{2\alpha^2},$$

$$\delta_x=\frac{A_{13}+2A_{55}-\alpha^2}{\alpha^2},\quad \delta_y=\frac{A_{23}+2A_{44}-\alpha^2}{\alpha^2},$$

$$\delta_z=\frac{A_{12}+2A_{66}-\alpha^2}{\alpha^2},$$

$$\chi_x=\frac{A_{14}+2A_{56}}{\alpha^2},\quad \chi_y=\frac{A_{25}+2A_{46}}{\alpha^2},\quad \chi_z=\frac{A_{36}+2A_{45}}{\alpha^2},$$

$$\epsilon_{15}=\frac{A_{15}}{\alpha^2},\quad \epsilon_{16}=\frac{A_{16}}{\alpha^2},\quad \epsilon_{24}=\frac{A_{24}}{\alpha^2},$$

$$\epsilon_{26}=\frac{A_{26}}{\alpha^2},\quad \epsilon_{34}=\frac{A_{34}}{\alpha^2},\quad \epsilon_{35}=\frac{A_{35}}{\alpha^2},$$

$$\epsilon_{46}=\frac{A_{46}}{\alpha^2},\quad \epsilon_{56}=\frac{A_{56}}{\alpha^2},\quad \epsilon_{45}=\frac{A_{45}}{\beta^2},$$

$$\gamma_x=\frac{A_{55}-\beta^2}{2\beta^2},\quad \gamma_y=\frac{A_{44}-\beta^2}{2\beta^2},\quad \gamma_z=\frac{A_{66}-\beta^2}{2\beta^2}.$$

(A3)

Although the WA parameters depend on velocities α and β of the reference isotropic medium, the matrix $\overline{B}_{mn}$ is *independent* of these velocities. Its elements have the following form:

$$\begin{aligned}
\overline{B}_{11}&=\beta^2+2\alpha^2D^{-2}(n_3^3[(\epsilon_{15}-\epsilon_{35})n_1^3+(\epsilon_{24}-\epsilon_{34})n_2^3+(\chi_x-\epsilon_{34})n_1^2n_2+(\chi_y-\epsilon_{35})n_1n_2^2]\\
&\quad+n_3^2[(\epsilon_x-\delta_x+\epsilon_z)n_1^4+(\epsilon_y-\delta_y+\epsilon_z)n_2^4+(\delta_z-\delta_x-\delta_y+2\epsilon_z)n_1^2n_2^2+2(\epsilon_{16}-\chi_z)n_1^3n_2+2(\epsilon_{26}-\chi_z)n_1n_2^3]\\
&\quad+n_3[(\epsilon_{35}-\epsilon_{15})n_1^5+(\epsilon_{34}-\epsilon_{24})n_2^5+(\epsilon_{34}-\chi_x)n_1^4n_2+(\epsilon_{35}-\chi_y)n_1n_2^4\\
&\quad+(2\epsilon_{35}-\chi_y-\epsilon_{15})n_1^3n_2^2+(2\epsilon_{34}-\chi_x-\epsilon_{24})n_1^2n_2^3])+2\beta^2D^{-2}(\gamma_xn_1^2+\gamma_yn_2^2+\epsilon_{45}n_1n_2),\\
\overline{B}_{12}&=\alpha^2D^{-2}(n_3^2[(\chi_x-\epsilon_{34})n_1^3-(\chi_y-\epsilon_{35})n_2^3+(\epsilon_{35}+\chi_y-2\epsilon_{15})n_1^2n_2+(2\epsilon_{24}-\chi_x-\epsilon_{34})n_1n_2^2]\\
&\quad+n_3[(\epsilon_{16}-\chi_z)n_1^4-(\epsilon_{26}-\chi_z)n_2^4+(\delta_x-\delta_y-2\epsilon_x+\delta_z)n_1^3n_2+(\delta_x-\delta_y+2\epsilon_y-\delta_z)n_1n_2^3+3(\epsilon_{26}-\epsilon_{16})n_1^2n_2^2]\\
&\quad+[(\epsilon_{15}-\chi_y)n_1^4n_2+(\chi_x-\epsilon_{24})n_1n_2^4+(\epsilon_{15}-\chi_y)n_1^2n_2^3+(\chi_x-\epsilon_{24})n_1^3n_2^2+\epsilon_{46}n_2^3-\epsilon_{56}n_1^3+\epsilon_{46}n_1^2n_2-\epsilon_{56}n_1n_2^2])\\
&\quad+\beta^2D^{-2}n_3[2(\gamma_y-\gamma_x)n_1n_2+\epsilon_{45}n_1^2-\epsilon_{45}n_2^2],\\
\overline{B}_{22}&=\beta^2+2\beta^2D^{-2}n_3^2(\gamma_xn_2^2+\gamma_yn_1^2-\epsilon_{45}n_1n_2)\\
&\quad+2\alpha^2D^{-2}(n_3[(\epsilon_{46}n_1^3+\epsilon_{56}n_2^3)+(\epsilon_{56}-\chi_x+\epsilon_{24})n_1^2n_2+(\epsilon_{46}-\chi_y+\epsilon_{15})n_1n_2^2]\\
&\quad+(\epsilon_x+\epsilon_y-\delta_z)n_1^2n_2^2+(\epsilon_{26}-\epsilon_{16})n_1^3n_2-(\epsilon_{26}-\epsilon_{16})n_1n_2^3])+2\beta^2\gamma_z(n_1^2+n_2^2),\\
\overline{B}_{13}&=\alpha^2D^{-1}\{n_3^4(\epsilon_{34}n_2+\epsilon_{35}n_1)+n_3^3[(\delta_x-\epsilon_x-\epsilon_z)n_1^2+(\delta_y-\epsilon_y-\epsilon_z)n_2^2+2\chi_zn_1n_2]\\
&\quad+n_3^2[(4\chi_x-3\epsilon_{34})n_1^2n_2+(4\chi_y-3\epsilon_{35})n_1n_2^2+(4\epsilon_{15}-3\epsilon_{35})n_1^3+(4\epsilon_{24}-3\epsilon_{34})n_2^3]\\
&\quad+n_3[(2\delta_z-\delta_x-\delta_y-\epsilon_x-\epsilon_y+2\epsilon_z)n_1^2n_2^2+2(2\epsilon_{16}-\chi_z)n_1^3n_2+2(2\epsilon_{26}-\chi_z)n_1n_2^3\\
&\quad+(\epsilon_x+\epsilon_z-\delta_x)n_1^4+(\epsilon_y+\epsilon_z-\delta_y)n_2^4+(\epsilon_x-\epsilon_z)n_1^2+(\epsilon_y-\epsilon_z)n_2^2]-\chi_xn_1^2n_2-\chi_yn_1n_2^2-\epsilon_{15}n_1^3-\epsilon_{24}n_2^3\},\\
\overline{B}_{23}&=\alpha^2D^{-1}\{n_3^3(\epsilon_{34}n_1-\epsilon_{35}n_2)+n_3^2[(\delta_y-\delta_x-\epsilon_y+\epsilon_x)n_1n_2+\chi_zn_1^2-\chi_zn_2^2]\\
&\quad+n_3[(2\chi_y-3\epsilon_{15})n_1^2n_2-(2\chi_x-3\epsilon_{24})n_1n_2^2+\chi_xn_1^3-\chi_yn_2^3]\\
&\quad+(\delta_z-\epsilon_x-\epsilon_y)n_1^3n_2-(\delta_z-\epsilon_x-\epsilon_y)n_1n_2^3+3(\epsilon_{26}-\epsilon_{16})n_1^2n_2^2+\epsilon_{16}n_1^4-\epsilon_{26}n_2^4+(\epsilon_y-\epsilon_x)n_1n_2\},\\
\overline{B}_{33}&=\alpha^2+2\alpha^2[2n_3^3(\epsilon_{34}n_2+\epsilon_{35}n_1)+n_3^2((\delta_x-\epsilon_x-\epsilon_z)n_1^2+(\delta_y-\epsilon_y-\epsilon_z)n_2^2+2\chi_zn_1n_2+\epsilon_z)\\
&\quad+2n_3(\chi_xn_1^2n_2+\chi_yn_1n_2^2+\epsilon_{15}n_1^3+\epsilon_{24}n_2^3)+\epsilon_xn_1^2+\epsilon_yn_2^2+(\delta_z-\epsilon_x-\epsilon_y)n_1^2n_2^2+2\epsilon_{16}n_1^3n_2+2\epsilon_{26}n_1n_2^3].
\end{aligned}$$

(A4)

Červený, V. (**2001**). *Seismic Ray Theory* (Cambridge U.P., Cambridge).
Chu, Y. C., Degtuar, A. D., and Rokhlin, S. I. (**1994**). ''On determination of orthotropic material moduli from ultrasonic velocity data in nonsymmetry planes,'' J. Acoust. Soc. Am. **95**, 3191–3203.
Farra, V. (**2001**). ''High order expressions of the phase velocity and polarization of qP and qS waves in anisotropic media,'' Geophys. J. Int. **147**, 93–105.
Helbig, K. (**1993**). ''Longitudinal directions in media of arbitrary anisotropy,'' Geophysics **58**, 680–691.
Jech, J., and Pšenčík, I. (**1989**). ''First-order perturbation method for anisotropic media,'' Geophys. J. Int. **99**, 369–376.
Pšenčík, I., and Gajewski, D. (**1998**). ''Polarization, phase velocity and NMO velocity of *qP* waves in arbitrary weakly anisotropic media,'' Geophysics **63**, 1754–1766.
Pšenčík, I., and Vavryčuk, V. (**2002**). ''Approximate relation between the ray vector and the wave normal in weakly anisotropic media,'' Stud. Geophys. Geod. **46**, 793–807.
Shearer, P. M., and Chapman, C. H. (**1989**). ''Ray tracing in azimuthally anisotropic media—I. Results for models of aligned cracks in the upper crust,'' Geophys. J. Int. **96**, 51–64.
Schoenberg, M., and Helbig, K. (**1997**). ''Orthorhombic media: Modeling elastic wave behavior in a vertically fractured earth,'' Geophysics **62**, 1954–1974.
Shuvalov, A. L., and Every, A. G. (**1997**). ''Shape of the acoustic slowness surface of anisotropic solids near points of conical degeneracy,'' J. Acoust. Soc. Am. **101**, 2381–2383.
Song, L.-P., Every, A. G., and Wright, C. (**2001**). ''Tomography of joint P-wave traveltime and polarization data: A simple approach for media with low to moderate velocity gradients,'' J. Acoust. Soc. Am. **109**, 1796–1807.
Vavryčuk, V. (**2003**). ''Parabolic lines and caustics in homogeneous weakly anisotropic solids,'' Geophys. J. Int. **152**, 318–334.
Zheng, X., and Pšenčík, I. (**2002**). ''Local determination of weak anisotropy parameters from qP-wave slowness and particle motion measurements,'' Pure Appl. Geophys. **159**, 1881–1905.

Infrasonic wind-noise reduction by barriers and spatial filters

Michael A. H. Hedlin[a)]
Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0225

Richard Raspet
National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677

(Received 2 October 2002; revised 10 June 2003; accepted 12 June 2003)

This paper reports experimental observations of wind speed and infrasonic noise reduction inside a wind barrier. The barrier is compared with ''rosette'' spatial filters and with a reference site that uses no noise reduction system. The barrier is investigated for use at International Monitoring System (IMS) infrasound array sites where spatially extensive noise-reducing systems cannot be used because of a shortage of suitable land. Wind speed inside a 2-m-high 50%-porous hexagonal barrier coated with a fine wire mesh is reduced from ambient levels by 90%. If the infrasound wind-noise level reductions are all plotted versus the reduced frequency given by f^*L/v, where L is the characteristic size of the array or barrier, f is the frequency, and v is the wind speed, the reductions at different wind speeds are observed to collapse into a single curve for each wind-noise reduction method. The reductions are minimal below a reduced frequency of 0.3 to 1, depending on the device, then spatial averaging over the turbulence structure leads to increased reduction. Above the reduced corner frequency, the barrier reduces infrasonic noise by up to 20 to 25 dB. Below the corner frequency the barrier displays a small reduction of about 4 dB. The rosettes display no reduction below the corner frequency. One other advantage of the wind barrier over rosette spatial filters is that the signal recorded inside the barrier enters the microbarometer from free air and is not integrated, possibly out of phase, after propagation through a system of narrow pipes. [DOI: 10.1121/1.1598198]

PACS numbers: 43.28.Dm, 43.28.We [LCS]

I. INTRODUCTION

A. Air flow and infrasonic noise

It is well known that a direct scaling relationship exists between air flow and acoustic noise in the atmosphere. The inverse relationship between pressure and velocity for a nonviscous fluid with steady flow is given by Bernoulli's principle

$$p+(\rho v^2)/2+\rho g h=C, \tag{1}$$

where p is pressure, ρ is density, v is wind velocity, g is the acceleration due to gravity, h is the height, and C is a constant. Velocity and pressure are inversely related if h is held constant and if energy is conserved. Differentiation of this formula, with ρ and h held constant, shows the simplicity of this scaling

$$\delta p \sim \rho v\, \delta v. \tag{2}$$

Morgan and Raspet (1992) measured the incoming wind speed and wind fluctuation on a microphone and demonstrated that the pressure fluctuations obey Eq. (2) within a constant factor ranging from 1.1 to 1.6. Much of the acoustic noise at short periods is due to turbules that are carried past the recording site, or advected, by ambient wind (Grover, 1971) with the frequency band of the noise governed largely by the scale of the turbules (McDonald, Douze, and Herrin, 1971) and, as we will show later, by the wind speed. Strasberg has discussed the physical mechanisms of infrasonic flow noise underwater as measured by hydrophones (Strasberg, 1979) and has applied this analysis to develop scaling laws for microphone wind screens in nonturbulent flows (Strasberg, 1988). Other sources of acoustic noise in the atmosphere are not directly related to the flow of air at the observation point, but are due to buoyancy waves and larger scale convective activity in the atmosphere (Priestley, 1966; Gossard and Hooke, 1975). The frequency dependence of this noise is governed by the spatial structure of turbulence in the atmosphere in the vicinity of the point where the noise is being measured. In general terms, atmospheric turbulence is concentrated near Earth's surface in the atmospheric boundary layer (ABL; Kaimal and Finnigan, 1994). The upper extent of the ABL is highest (1 to 2 km) during the day when the Earth's surface radiates heat and lowest (tens of meters) or nonexistent during the night (Panofsky and Dutton, 1984; Kaimal and Finnigan, 1994). Wind shear introduces smaller eddies and ''cascades'' energy from the low to the higher frequencies (Kaimal and Finnigan, 1994).

B. The new global infrasound monitoring network

The tabling of the Comprehensive nuclear Test-Ban Treaty (CTBT) at the United Nations in September, 1996 increased interest in monitoring globally for nuclear tests at all explosive yields. The treaty comes with an International Monitoring System (IMS), which will include networks of seismic, infrasonic, hydro-acoustic, and radionuclide stations. This paper relates to the infrasound component of the

a)Electronic mail: hedlin@ucsd.edu

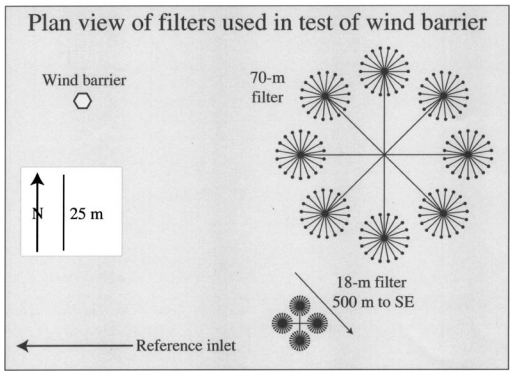


FIG. 1. Plan view of all filters tested in this experiment. The 70- and 18-m rosette filters are part of the International Monitoring System (IMS) array I57US. The 18-m filter is located 500 m to the southeast of the main study area. The large rosette filter comprises 144 low-impedance inlets distributed across an area 70 m in diameter. The small rosette filter comprises 96 low-impedance inlets distributed across an area 18 m in diameter. These filters are described in more detail in Alcoverro (1998) and Hedlin *et al.* (2003). All filters are shown to scale.

IMS. The infrasound network will comprise 60 arrays distributed globally. Each array will include up to eight sensors. The dynamic atmosphere is inherently noisy at frequencies of interest to the nuclear monitoring community (between 0.01 and 8 Hz). As a result, it is necessary to suppress ambient infrasonic noise at each station in the infrasound network. IMS stations are currently being equipped with ''rosette'' spatial noise reduction systems (Fig. 1; Alcoverro, 1998; Hedlin, Alcoverro, and D'Spain, 2003). To achieve a spatially uniform global network, 23 of the stations are located on oceanic islands.

C. Rationale for the present study

As development of the IMS global infrasound network progresses, and given that little suitable land is available at some of the sites, there is a pressing need for effective, spatially compact, passive infrasonic noise-reducing devices. An intercomparison of the noise-reducing systems currently deployed at new IMS infrasound array sites and spatially compact systems, such as the wind barrier, has not yet been done. Grover (1971) compared very small wind screens (<1 m high) with pipe filters that are not still in use. This paper reports an intercomparison of a small, 2-meter-high, wind- and noise-reducing barrier with large spatial filters that are currently in use at IMS infrasound array sites. Noise levels inside the barrier are also compared with ambient noise levels to gauge, in an absolute sense, the utility of the barrier for reducing infrasonic noise across the frequency band of interest to the nuclear monitoring community. The physics underlying interest in the wind barrier is straightforward. Noise scales with wind, and therefore reducing wind, and fragmenting advected turbules, at the site of the microbarometer should lead to lower noise levels and an improved ratio of signal to noise. The microbarometer we used in our experiment is described later in the paper (Sec. II E). This paper seeks answers to three questions: (1) What is the frequency band over which a wind barrier reduces noise? (2) By how much is the noise reduced in this band? And (3) Does the barrier compare favorably with spatial filters as a system for preserving signals of interest while attenuating noise?

II. EXPERIMENT LAYOUT

A. The infrasound test-bed

The experiment was conducted at the Cecil H. and Ida M. Green Pinon Flat Observatory in the Anza-Borrego desert in southern California during a 40-day interval in August and September of 2000. The Pinon Flat Observatory is useful for studies of infrasonic noise-reducing systems as wind speeds in this area range from near zero at night to somewhat above 10 m/s during the day. In addition to strong diurnal variations in meteorological conditions, the area experiences significant seasonal variations. The infrasonic noise levels scale directly with wind speed (McDonald, Douze, and Herrin, 1971; Hedlin, Alcoverro, and D'Spain, 2003). A diagram showing the design of the spatial filters we compared with the wind barrier and the relative locations of all filters used in our study is given in Fig. 1.

B. Design considerations for the wind barrier

The design of the wind barrier used in this experiment was guided partly by previous work by Lizska, and by Bernoulli's scaling between wind speed and pressure. The barrier chosen for this site is 2 m tall and 5.5 m across at the base. To avoid increasing large-scale turbulence downstream of the barrier, we followed Lizska and used a barrier that was 50% porous on the sides. As shown in Fig. 2, the openings in the hexagonal barrier are horizontal. The barrier was coated with a fine mesh to further decrease wind flow inside. The wire mesh used had square openings 1 mm on a side. As we see later, the observed reduction in infrasonic noise does not scale as expected with the reduction in wind speed, pointing to a different mechanism for noise reduction inside the barrier that relies on averaging pressure variations across the surface of the barrier and not directly on the reduction of wind speed at the microbarometer inside the barrier.

To test the wind barrier, infrasonic noise and meteorological data were collected at four sites. A microbarometer and a wind sensor were located inside the hexagonal wind barrier (Fig. 2) at a height of 60 cm. The other sensors were located at the sites described in the following sections.

C. Rosette spatial filters

The second and third sensors were attached to 18- and 70-m-diameter ''rosette'' spatial noise-reducing filters at the nearby IMS infrasound array I57US (Fig. 1). The rosette filters comprise a number of low-impedance inlets distributed across a circular area (Alcoverro, 1998; Hedlin, Alcoverro, and D'Spain, 2003). Atmospheric noise from local sources is known to be incoherent at offsets of several meters or less, while signals from remote sources can be coherent at distances of hundreds of meters (Priestley, 1966). Pressure variations sampled by N inlets are summed, with signal-to-noise levels improved by $N^{1/2}$ at sites where the noise is uncorrelated between the inlets (Daniels, 1959). The rosettes provide spatial averaging over the local turbulent structures.

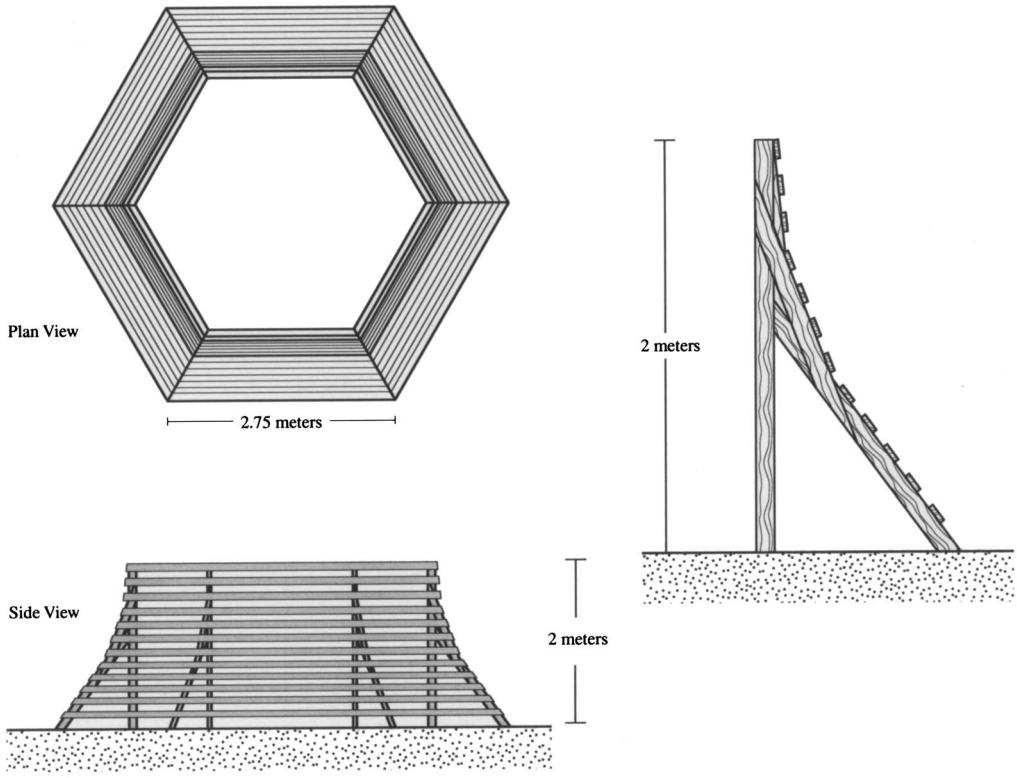


FIG. 2. The wind barrier considered in this paper consists of six sides, each 2 m tall. The barrier is not solid on the sides, as theory indicates this would increase turbulence and infrasonic noise. The sides of the barrier are 50% porous (i.e., one half solid). Additional high-frequency noise reduction was achieved by coating the sides of the barrier with a fine wire mesh similar to residential bug screens. The top of the barrier was just coated with the wire mesh but otherwise uncovered. The design of our barrier was suggested by Ludwik Liszka. Exact specifications were given by Doug Revelle (Los Alamos National Laboratory). The MB2000 sensor was located in foam within a porous aluminum shell at an elevation of 60 cm at the center of the barrier. The foam provided additional reduction of high-frequency noise; the shell protected the sensor from the elements. This barrier was patented by Liszka in 1975 (Liszka, L.: Swedish Patent No. 7315138-3, October 30, 1975). The drawing in this figure was taken from this patent document.

D. Reference noise measurements

To allow us to gauge the utility of the wind barrier at reducing infrasonic noise from ambient levels, all recordings inside the wind barrier were made concurrently with recordings made at the fourth site equipped with no noise-reducing system. At the reference site, one sensor was attached to a single, low-impedance inlet located 5 cm above the ground. See Fig. 3.

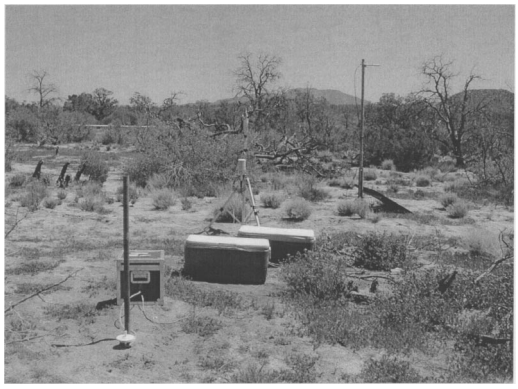


FIG. 3. Photo of reference site. The low-impedance inlet is located 5 cm above the ground. The microbarometer is located in a case immediately behind the inlet. The recording equipment and power system are located to the right of the microbarometer. The wind and temperature sensors are located 2 m and 1 m above the ground immediately behind the recording equipment. The radio antenna is located to the far right.

E. Recording system

Infrasonic pressure data were collected at all four sites located within an area 150 m across. No sites were located downwind from the barrier to avoid artificial sheltering from the wind. All pressure data were collected with MB2000 aneroid microbarometers. The sensors were fabricated by the French Département Analyse et Surveillance de l'Environnement (DASE). The MB2000 sensor filters the signal between 0.01 and 27 Hz. The recording system filters the data below 9 Hz to avoid aliasing. Pressure data were sampled at all sites at 20 Hz. Meteorological data (wind speed, direction, air temperature, and humidity) were sampled at all four sites at 1 Hz. All data were digitized by 24-bit Reftek dataloggers and transmitted in real time via 2.2-GHz telemetry link to our laboratory in La Jolla. Power for all equipment was provided via solar panels and batteries.

III. EXPERIMENTAL RESULTS

The core of the observational work has been the spectral analysis of segments of infrasonic and meteorological data that were collected simultaneously by identical microbarometers attached to different noise-reduction systems—or to no noise-reduction device at all. The data analysis is based on 15-min segments of data taken from nonoverlapping intervals throughout the experiment. Every hour, four segments of data were selected for further analysis. A segment was

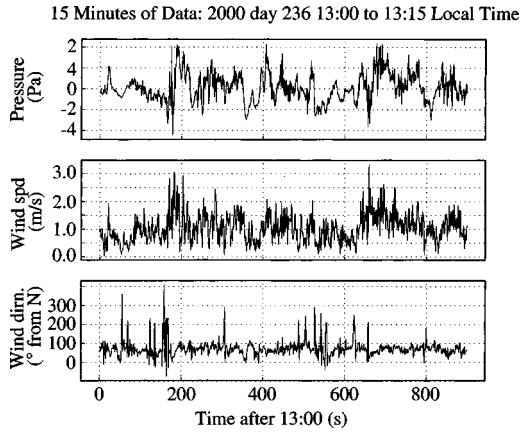


FIG. 4. Atmospheric pressure and meteorological data from a 15-min interval starting at 13:00 GMT on day 236 of 2000 at the reference site. Filtered pressure data are shown in the upper panel. Wind speed and direction are shown below.

removed from consideration only if data were lost during that time interval at any of the four sites. The results presented in this paper came from 40 days of observations and involve a total of 3074 spectral estimates and over 3 million observations of wind velocity at each site. One example of infrasonic pressure and meteorological data is given in Fig. 4. This figure presents a typical 15-min segment of data collected inside the wind barrier. During the entire experiment, the wind speed varied from near zero to above 10.0 m/s.

A. Wind-speed reduction

Figure 5 shows the effectiveness of the wind barrier at reducing the wind speed. The test of wind reduction occurred in two phases. In both phases, wind speed was measured at the center of the barrier at a height of 60 cm and outside at a height of 2 m at the location of the reference inlet. In the first phase, a 50% porous wind barrier was used. In the second phase, the sides and top of the barrier were coated with a fine wire mesh. The mesh is composed of a Cartesian grid of wires spaced by 1 mm. The histograms in the lower panel of Fig. 5 show the number of wind-speed observations made as a function of ambient wind speed (i.e., wind speed recorded outside the barrier at the reference site). To calculate each histogram, the wind-speed data at the reference site were binned into intervals spanning 0.1 m/s. Most observations from both phases of the experiment were made at wind speeds below 2.0 m/s. The winds were slightly stronger during the phase of the experiment when the barrier was coated with a mesh. The upper panel displays the average wind speed observed inside the barrier and the ambient wind speed recorded at the reference site. To calculate the average as a function of wind speed, individual observations were binned as described above. Each average was calculated from simultaneous observations made inside the barrier and at the reference site. A barrier that provided no shielding from the wind would yield a curve lying along the dashed line. The data show that the uncoated barrier reduces the wind speed by ~80% at times when the unsuppressed wind speed exceeded 10 m/s. The addition of the wire mesh increased the wind reduction to 90%.

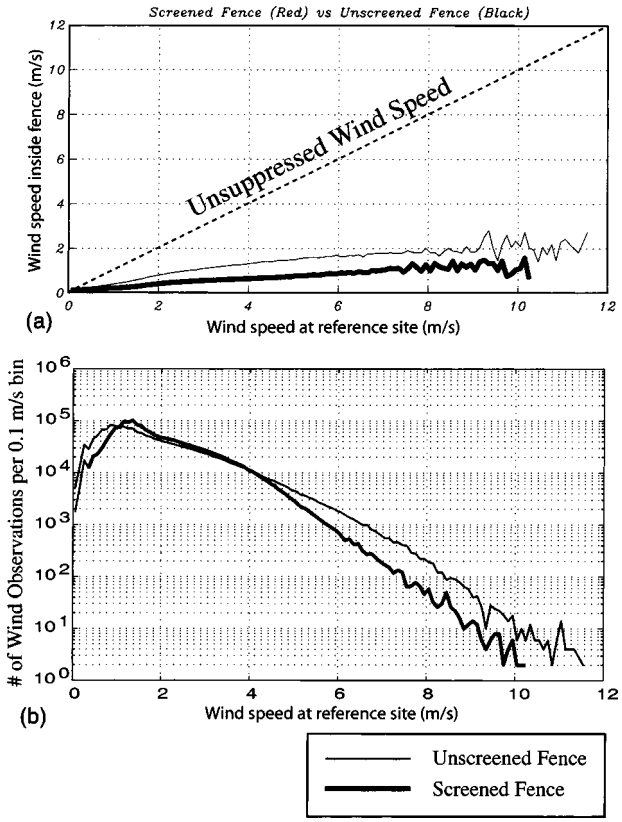


FIG. 5. Data collected by anemometers at a height of 60 cm within the wind barrier and outside at a height of 2 m indicates that the barrier is highly effective at reducing wind speed. The lower panel shows the number of observations at the reference site as a function of wind speed. This figure is based on a total of 3.3 million wind-speed observations (made once/second). The faint curve represents data collected before the wind barrier was coated with a wire mesh. The dark curve represents data collected after the sides of the wind barrier were coated with the mesh. The upper panel compares the wind speed inside the barrier to that observed outside. If the wind barrier was completely ineffective at reducing wind speed, the data would lie along the dashed line. The 50%, uncoated, wind barrier reduces wind speed by ~80%. The addition of the mesh reduces wind speed further to 90%. As shown in panel (b), the ambient wind speeds were slighly higher during the test of the barrier with the external mesh. It must also be noted that because the wind sensor inside the barrier was closer to the ground than the sensor outside at the reference site (60 cm vs 2.0 m), some wind-speed reduction would be expected even in the absence of a wind barrier. A separate experiment indicated that average wind speeds at the observatory at a height of 60 cm are 60% of the wind speeds at 2.0 m.

B. Noise reduction

Although the wind barrier provided substantial shielding from the wind, the main objective of this paper is to judge the effectiveness of the barrier at reducing infrasonic noise. Figure 6 summarizes this result. Given that individual power spectral estimates are inherently noisy, the analysis is based on stacked spectra. In this analysis, each of the 3074 spectral estimates is associated with a single wind-speed value and is binned by that value. The wind-speed value associated with an individual power spectral estimate is the average wind speed at the reference site from the 15-min time interval spanned by the spectral estimate. The wind-speed bins used in this analysis spanned 0.5 m/s, starting at 0.0 m/s. Each panel in Fig. 6 displays stacked power spectral estimates from each of the four sites. For example, the upper-left panel shows in black the average ambient noise levels when the

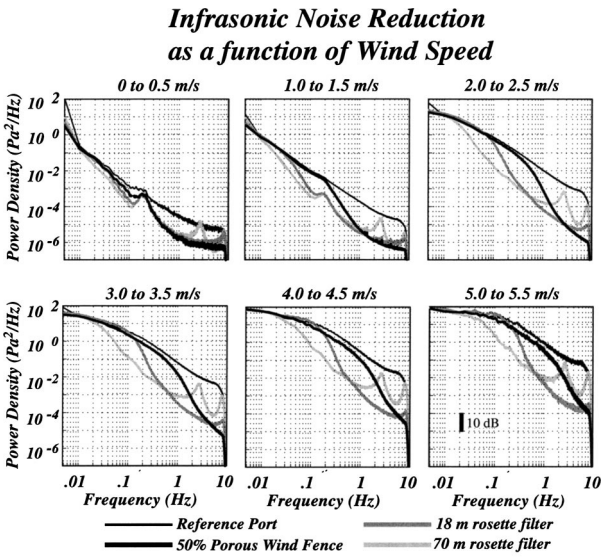


FIG. 6. Due to high variance in individual power spectral estimates, individual estimates are grouped by wind speed and stacked to provide a clearer view of the utility of the wind barrier at reducing infrasonic noise and the dependence of infrasonic noise as a function of wind speed and frequency. Stacked power spectral density estimates taken from data collected at four sites over a 40-day interval are shown at wind speeds ranging from <0.5 m/s to above 5.0 m/s. Ambient noise levels are indicated by the black curves in each panel. Data collected inside the 50%-porous, screened, wind barrier are represented by the thick black curves. For reference, data collected via ''rosette'' filters at a nearby IMS infrasound array (I57US) are also shown. Data collected via a 70-m aperture rosette filter are shown in light gray. The thin black curves represent data collected via a single reference inlet. Data from a 18-m rosette filter are shown in heavy gray. The rosette filters are described in more detail by Alcoverro (1998) and Hedlin, Alcoverro, and D'Spain (2003). The stacked spectra reveal the increase in infrasonic noise with decreasing frequency at all wind speeds. Noise levels increase with increasing wind speed at all frequencies; however, the greatest increase in noise levels occurs at high frequencies. The screened wind barrier provides noise suppression at high frequencies. The corner frequency of the filter increases with increasing wind speed from $\sim$0.2 Hz at wind speeds below 0.5 m/s to $\sim$1.0 Hz at wind speeds above 5.0 m/s. All wind data used in this figure were collected at a height of 2.0 m at the reference inlet. In total, 3074 15-min intervals of data were used in this figure. Each interval provided a single power spectral density estimate and a single wind-speed value determined by taking an average of all wind-speed measurements made during the 15 min covered by the spectral estimate.

wind speed at the reference site was below 0.5 m/s. The noise levels observed in the barrier, at the 70-m rosette and the 18-m rosette filter are shown in black, light-, and dark gray, respectively. This figure shows the growth of infrasonic noise at all sites with increasing wind speed and with decreasing frequency. More importantly for this study, this figure shows the relative utility of the different systems.

The frequency band over which the barrier is effective scales with wind speed. At high frequencies, the screened 50%-porous wind barrier reduces infrasonic noise by 20 to 25 dB and is superior to the spatial filters at frequencies above 1 to 5 Hz, with the greatest advantage observed at low wind speeds. The wind barrier is less effective than the spatial filters at reducing infrasonic noise at all wind speeds at frequencies below 1–2 Hz. The barrier offers 0- to 5-dB noise reduction at frequencies below 1 Hz. Resonance is observed in data from the rosette filter at frequencies above 0.7 Hz (Fig. 6; Hedlin, Alcoverro, and D'Spain, 2003). The microbarom peak is observed inside the wind barrier only under calm conditions.

IV. FREQUENCY DEPENDENCE OF THE WIND NOISE REDUCTION

Figure 7 displays the wind-noise reduction versus scaled frequency fL/v, where f is the frequency, L the device scale size, and ν the average wind speed for the three highest wind speeds. In this plot, the diameters of the rosettes and wind barrier are used (18, 70, and 5.5 m). The diameter of the wind barrier is the distance at ground level between opposite sides. Each set of data is offset by a factor of 100.

If the wind noise and wind-noise reduction are caused by local interactions of the turbulence and the wind-noise reduction device, the wind-noise reductions of similar devices should scale as the ratio of the turbulence scale to a characteristic linear dimension of the device (Strasberg, 1988). The turbulence size corresponding to a given frequency is given by the average wind speed divided by the frequency. Thus, Fig. 7 displays the relative effectiveness of a device weighted by its size.

The behavior of the reduction versus scaled frequency for each wind speed is very similar. The contribution from the resonance in the 70-m rosette does not scale with wind speed, but the other features do. The two rosettes are effective at spatially averaging the turbulent pressure fluctuations if the scaled frequency is greater than 1. There are minor differences between the 18- and 70-m rosette, but the scaling appears to hold well in general. The wind barrier appears to roll off at a higher scaled frequency, indicating that the spatial averaging is occurring over a smaller region. Reasonable agreement between the roll-off frequency of the rosettes and the wind barrier can be obtained by using the barrier height as the averaging scale. This is presented in Fig. 8 along with additional data from spherical wind screens which will be discussed in Sec. V. In Fig. 8 it can be observed that the wind barrier produces a small but significant 4-dB reduction for scaled frequencies between 0.05 and 0.5.

V. DISCUSSION

The data presented above display several unexpected results. A reduction in wind speed inside the barrier by a factor of 10 does not result in a proportional reduction in infrasonic wind-noise power by a factor of 100. Second, the wind-noise reduction by the barrier scales much like the wind-noise reduction of the rosettes produced by spatial averaging. The appropriate size for scaling of the barrier appears closer to the height of the barrier, 2.0 m, rather than the diameter, 5.5 m. The barrier does display a small but significant reduction even when the size of the turbulence is greater than the size of the barrier itself. On the scaled plot, assuming the effective size of the barrier is 2 m, the infrasonic spectrum is reduced by a factor of 1/(2.5), or by -4 dB over the scaled frequency range of 0.05 to 0.5.

Some insight into the mechanism for this reduction can be gained by a study of the literature of wind-noise reduction at audio frequencies. Phelps (1938) suggested that the mechanism for wind-noise reduction for low frequencies is

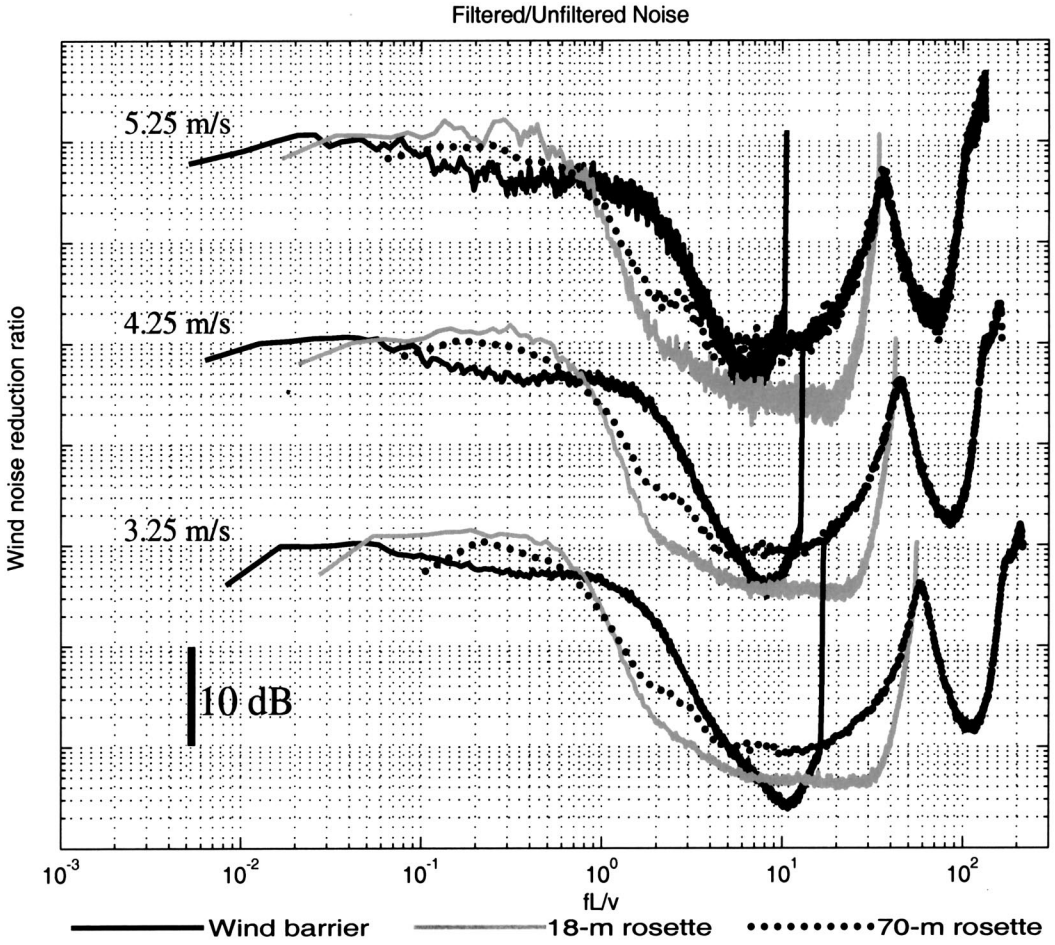


FIG. 7. Wind-noise reduction versus scaled frequency at 3.25, 4.25, and 5.25 m/s for the 70-m rosette (bold dots), the 18-m rosette (bold gray), and the 5.5-m wind barrier (black curves).

as follows: the pressure profile induced by the wind over a spherical screen has regions of high and low pressure. Since the pressure fluctuations in incompressible flows are governed by the Laplace equation in regions of little or no flow (inside the screen), the pressure at the microphone can be calculated by area averaging over the screen surface. Phelps used the equations for the pressure distribution for steady inviscid incompressible flow over a rigid sphere to demonstrate a predicted reduction of 6.0 dB for turbulence scales larger than the size of the windscreen. The pressure distribution around a rigid sphere is dominated by a strong low-pressure region behind the wind screen. Morgan measured the pressure distribution around 90- and 180-mm diameter spherical foam windscreens and used the measured data to evaluate the reduction (Morgan, 1992). He found that the small flow through the porous windscreen leads to a smaller low-pressure region behind the wind screen and therefore a smaller average pressure. He arrived at a value of 14 dB (0.04 relative power) for this reduction. Data from Morgan's dissertation for the wind-noise reduction of a 90- and 180-mm screen for 4.84- and 4.74-m/s wind speeds in the frequency bands from 1.6 to 500 Hz are also displayed in Fig. 8, where the characteristic length dimension corresponds to the diameter of the spherical windscreen. At small scale sizes, the data agree reasonably well with the prediction of 14 dB. The reductions are not constant but vary slowly up to scale screen sizes of 0.3. As the scale size increases beyond 0.3, additional reduction occurs since different portions of the screen are now sensing different turbulent structures. This averaging is comparable to the averaging by the rosettes. The reduction in the low-pressure region behind the screen due to flow-through also reduces the intensity of wake turbulence and displaces the onset of turbulence further from the screen (Schomer *et al.*, 1990).

It appears that the wind barrier may be acting somewhat like a foam windscreen in that it does display reductions in wind-noise intensity even when the turbulence structures are larger than the screen. The reductions are not as large as displayed for the spherical foam windscreens since the top of the barrier is open. In addition, the pressure distribution is not known and may be far from optimal since the pressure-averaging mechanism was not considered in the barrier design.

VI. CONCLUDING REMARKS

A. Wind-noise reduction

Comparison of the scaled reductions in wind noise produced by the rosettes and wind barrier with the reductions afforded by a spherical windscreen hold promise for significant wind-noise reduction with a smaller footprint device. The rosettes only produce reductions if the scale size of the

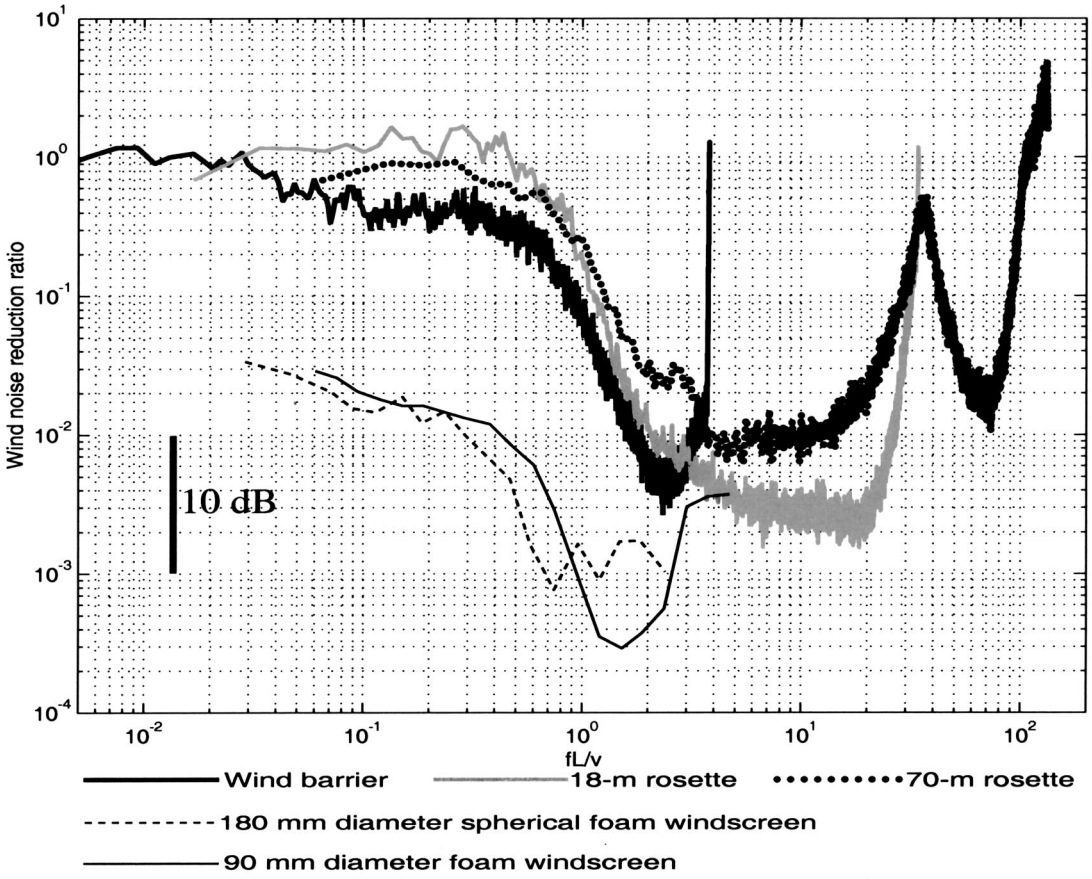


FIG. 8. Wind-noise reduction versus scaled frequency at 5.25 m/s for the 70-m rosette (bold dots), for the 18-m rosette (bold gray). and for the wind barrier (bold black) using the height of 2.0 m to calculate the scaled frequency. Also displayed is the noise reduction produced by a 90-mm-diameter foam windscreen at an average wind speed of 4.84 m/s (light dashed curve) and a 180-mm diameter spherical foam wind screen at 4.74 m/s (light solid curve).

turbulence is smaller than the size of the rosette, since such devices rely on the incoherence of the turbulence at each inlet. The wind barrier displayed large reductions only when the scale size of the turbulence is smaller than the height of the barrier. However, a reduction of about 4.0 dB was achieved by the wind barrier for turbulence scales up to 20 times the barrier height (reduced frequencies down to 0.05). This reduction may be due to a mechanism similar to that of the spherical foam windscreens. In the spherical windscreens, these reductions occur since the pressure measured at the center is the area average of the pressures generated at the surface of the sphere. For large turbules, the pressure generated by an increase in wind speed is positive at the front and negative at the back, and the average is less than the pressure fluctuation measured at a bare sensor. The wind barrier may serve as a pressure averaging device over the surface of the barrier with negative and positive contributions even for turbulent structures with dimensions much larger than the windscreen.

The similarity of the results of the wind barrier and the spherical foam ball suggest that a wind-barrier design based on the principle of spherical windscreen reduction may produce a more effective, more compact infrasonic wind-noise reduction device.

B. Preserving signals while attenuating noise

Ultimately, we seek a device that will provide maximum attenuation of unwanted noise and minimally distorted recordings of signals from remote sources. Signals and noise are received at a single point inside the barrier. This is potentially a significant advantage of the barrier over spatial filters, such as the two considered in this paper, which rely on the different coherence lengths of signal and noise. Spatial filters are used to increase the ratio of signal to noise by sampling air pressure at numerous locations at offsets at which the noise is believed to be incoherent and the signal remains coherent. If phasing between the multiple samples of the signal is not correctly taken into account, and if the signal is not coherent across the area spanned by the spatial filter, the waveform of the signal will be degraded. Rosette filters are tuned to signals arriving with infinite phase velocity. Signal attenuation caused by 70-m aperture rosette filters becomes acute at low angles of incidence and at frequencies above 1 Hz (Hedlin, Alcoverro, and D'Spain, 2003).

Unlike spatial filters currently in use at IMS array sites, the barrier does not propagate the signals through narrow pipes. Ambient signal and noise enter the microbarometer from free air and therefore, dispersion of broadband signals that is known to occur in narrow pipes (e.g., Benade, 1968) is not a concern.

ACKNOWLEDGMENTS

Pioneering work on the wind barrier by Ludwik Liszka (Swedish Institute of Space Physics) made this research possible. The design of the wind barrier and insights into how

the barrier reduces infrasonic noise were provided by Doug Revelle (Los Alamos National Laboratory). David Revelle constructed the barrier. We thank Hank Bass (University of Mississippi), Lou Sutherland (the associate editor), and two anonymous reviewers for constructive comments on the manuscript. Frank Vernon, Jennifer Eakins, and Glen Offield provided the real-time data link. Clint Coon provided field assistance. Funding was provided by the Defense Threat Reduction Agency under Contract No. DTRA01-00-C-0085.

Alcoverro, B. (**1998**). ''Acoustic filters design and experimental results,'' in Proceedings: Workshop on Infrasound, Commissariat à l'Energie Atomique, Bruyères-le-Châtel, France, 21–24 July 1998.

Benade, A. H. (**1968**). ''On the propagation of sound waves in a cylindrical conduit,'' J. Acoust. Soc. Am. **44**, 616–623.

Burridge, R. (**1971**). ''The acoustics of pipe arrays,'' Geophys. J. R. Astron. Soc. **26**, 53–69.

Daniels, F. P. (**1959**). ''Noise reducing line microphone for frequencies below 1 cps,'' J. Acoust. Soc. Am. **31**, 529–531.

Grover, F. H. (**1971**). ''Experimental noise reducers for an active microbarograph array,'' Geophys. J. R. Astron. Soc. **26**, 41–52.

Gossard, E. E., and Hooke, W. H. (**1975**). *Waves in the Atmosphere* (Elsevier Scientific, New York).

Hedlin, M. A. H., Berger, J., and Vernon, F. L. (**2002**). ''Surveying infrasonic noise on oceanic islands,'' Pure Appl. Geophys. **159**, No. 5, 1127–1152.

Hedlin, M. A. H., Alcoverro, B., and D'Spain, G. (**2003**). ''Evaluation of rosette infrasonic noise-reducing spatial filters,'' J. Acoust. Soc. Am. (to be published).

Kaimal, J. C., and Finnigan, J. J. (**1994**). *Atmospheric Boundary Layer Flows: Their Structure and Measurement* (Oxford University Press, Oxford).

McDonald, J. A., Douze, E. J., and Herrin, E. (**1971**). ''Structure of Atmospheric Turbulence,'' Geophys. J. R. Astron. Soc. **26**, 99–109.

Morgan, S., and Raspet, R. (**1992**). ''Investigation of the mechanisms of low-frequency wind noise generation outdoors,'' J. Acoust. Soc. Am. **92**, 1180–1183.

Morgan, M. S. (**1992**). ''An Investigation of the Sources and Attenuation of Wind Noise in Measurement Microphones,'' Ph.D. dissertation, University of Mississippi.

Panofsky, H. A., and Dutton, J. A. (**1984**). *Atmospheric Turbulence: Models and Methods for Engineering Applications* (Wiley, New York).

Phelps, W. (**1938**). ''Microphone wind screening,'' RCA Rev. **3**, 203–212.

Priestley, J. T. (**1966**). ''Correlation studies of pressure fluctuations on the ground beneath a turbulent boundary layer,'' NBS Report 8942, U.S. Dept. of Commerce, National Bureau of Standards.

Schomer, P. D., Raspet, R., Brunner, J., Marshal, D., Wagner, M., and Walker, D. (**1990**). ''Reduction of wind noise for unattended blast noise monitoring,'' Noise Control Eng. J. **34**(2), 77–88.

Strasberg, M. (**1979**). ''Nonacoustic noise interference in measurements of infrasonic ambient noise,'' J. Acoust. Soc. Am. **66**, 1487–1493.

Strasberg, M. (**1988**). ''Dimensional analysis of windscreen noise,'' J. Acoust. Soc. Am. **83**, 544–548.

Doppler sidebands in the cross-spectral density of narrow-band reverberation from a dynamic sea surface

Robert F. Gragg[a)]
Naval Research Laboratory, Washington, DC 20375-5350

(Received 12 December 2001; revised 7 April 2003; accepted 29 May 2003)

Analytic methods are used to formulate the impact of a random dynamic sea surface on the space-frequency characteristics of bistatic reverberation. A narrow-band point source is positioned beneath the time-dependent surface of a range-independent ocean. The small-waveheight perturbative approximation is invoked, and attention is focused on the Doppler sideband contributions to the reverberation cross-spectral density for an arbitrarily placed receiver pair. The new expression that results is identified as an active scattering generalization of the van Cittert–Zernike theorem from classical partial coherence theory. This work is the first to explicitly predict the sideband structure in the cross-spectral density of the field scattered from a realistic moving sea surface. A numerical example is presented for a shallow source and shallow receivers in a homogeneous ocean. [DOI: 10.1121/1.1595654]

PACS numbers: 43.30.Re, 43.30.Hw, 43.30.Vh [DLB]

I. INTRODUCTION

Although surface scattering can have several origins, this article deals only with the contribution due to the rough air/sea interface itself. This component typically dominates at low frequencies and/or wind speeds,[1] and at large scattering angles (such as the bistatic geometries considered here)[2] it may remain dominant even beyond that regime. The object of the effort is to determine the reverberant field's cross-spectral density (CSD); i.e., its second-order space-frequency statistics. The measured phone-to-phone spatial coherence of the acoustic field (i.e., the measured CSD matrix) is the generic input for a wide class of beamforming algorithms used with hydrophone arrays.[3,4] Algorithm development, however, is very often pursued using simulated CSDs. This article concentrates on producing a Doppler-sensitive formulation that can be used to simulate the sideband components of the reverberation CSD—a capability that may be useful in the development of signal-processing algorithms to distinguish reverberation from the echo of a moving target.

Perturbative approaches to air–sea interface scattering are fairly common in the literature, and have been reasonably successful at describing acoustic scattering effects.[5] Two deserve particular mention here. In the 1970s, Harper and Labianca published a series of articles that applied analytic techniques to sea-surface scattering.[6–10] They modeled the fundamental stochastic process (the surface wave field) in detail and developed expressions for the Doppler effects that it induces in the acoustic signal scattered to a bistatically placed hydrophone; i.e., in the power spectral density (PSD) of the reverberant acoustic field. In 1993, Goalwin addressed this general type of scattering problem at the level of second-order statistics,[11,12] extending the treatment to deal with the CSD for a pair of hydrophones. However, having assumed a ''frozen'' (motionless) sea surface, he could not address Doppler phenomena. The present article can be viewed as an extension of Goalwin's work into the realm of Doppler effects.

The problem is formulated below with a fixed narrow-band point source, a range-independent ocean, and a random dynamic sea surface. Analytic techniques are used to express the Doppler sidebands of the reverberation's CSD in terms of (1) the statistical descriptors of the sea surface (i.e., the CSD of the surface or, ultimately, its power spectrum and azimuthal directionality) and (2) the Green's function for the same environment without surface waves. The notation is outlined in Sec. II. Section III deals on a deterministic level with the two basic elements of the problem—the moving sea surface and the acoustic field that scatters from it. The makeup of a dynamic surface-wave field is discussed, and the small waveheight approximation (SWHA) is invoked to characterize the narrow-band scattered field. In Sec. IV, the surface elevation is treated as a random process that is stationary in time and in horizontal coordinates. An SWHA series is then obtained for the sideband components of the reverberation field's CSD. This expression is recognized as an active-scattering extension of the van Cittert–Zernike theorem from the classical theory of partially coherent wave fields. With the sea surface described in terms of its dispersion relation, power spectrum, and azimuthal directionality, each sideband is expressed as an azimuth integration involving a fundamental 2D wave number integral (the ''propagation integral''). This propagation integral is evaluated and an analytic expression is given for the Doppler sidebands of the reverberation CSD in the simple case of a uniform ocean with a shallow source and shallow receivers. In Sec. V, an example is presented to illustrate that case. Section VI provides a summary.

II. NOTATION

Vectors in two and three dimensions are indicated by boldface and arrows, respectively; e.g., $\vec{r}=(\mathbf{r},z)$. The same symbol in regular font is the magnitude of a $2D$ vector; e.g., r denotes $|\mathbf{r}|$, *not* $|\vec{r}|$. Directions in the horizontal plane are

[a)]Electronic mail: robert.gragg@nrl.navy.mil

often represented by a unit vector $\mathbf{n}(\theta)$ that points along the azimuth θ. The frequency f and its angular counterpart $\omega = 2\pi f$ both appear. The same is true of the horizontal wave vector $\mathbf{k} = 2\pi\mathbf{s}$.

The time and frequency domains are connected by the standard Fourier transformation, $a(f) = \int_{-\infty}^{+\infty} dt\, E(+ft)a(t)$, with the kernel $E(q) = \exp(i2\pi q)$. The same basic symbol is used in both domains—$a(t)$, $a(f)$—with only the argument indicating which domain is intended. This transform relation is often denoted $a(t) \xrightarrow{t \quad f} a(f)$. The 2D spatial transformation follows the same pattern, $a(\mathbf{r}) \xrightarrow{\mathbf{r} \quad \mathbf{s}} a(\mathbf{s}) = \int d^2\mathbf{r} E(-\mathbf{s}\cdot\mathbf{r})a(\mathbf{r})$, except that its kernel uses the opposite sign.

When $a(t)$ is effectively constant, the notation $a(\not t)$ is used to emphasize that it is virtually a dc signal. To the same approximation, the signal's spectrum is singular: $a(f) \approx a(f \sim 0)\delta(f)$, where $a(f \sim 0) = \int_{0-}^{0+} df a(f)$ denotes the strength of the singularity. Numerically, of course, $a(\not t) = a(f \sim 0)$. Similar notation is applied for spatial coordinates: when $a(\mathbf{r})$ is effectively independent of $\mathbf{r}$, it is denoted $a(\not\mathbf{r})$ and its wave number spectrum is written $a(\mathbf{s}) \approx a(\mathbf{s} \sim \mathbf{0})\delta(\mathbf{s})$, where $a(\mathbf{s} \sim \mathbf{0}) = a(\not\mathbf{r})$. These should all be understood in the same spirit as expressions such as "$x \ll y$"—as a form of physical notation rather than a purely mathematical one.

Fourier transformation can re-express an arbitrary field $a(\mathbf{r},t)$, a function of position and time, in terms of a dependence on wave number and frequency: $a(\mathbf{r},t) \xrightarrow{t \quad f} \xrightarrow{\mathbf{r} \quad \mathbf{s}} a(\mathbf{s},f)$, with the symmetry $a(\mathbf{s},f) = a^*(-\mathbf{s},-f)$ corresponding to a real-valued physical field [i.e., $a(\mathbf{r},t) = a^*(\mathbf{r},t)$]. It is often convenient to use complex-valued representations—$\alpha(\mathbf{r},t)$, with $\mathrm{Re}[\alpha(\mathbf{r},t)) = a(\mathbf{r},t)$]—though care must be used with nonlinear forms, notably quadratic field expressions such as $\alpha^*(\mathbf{r}_1,t_1)\alpha(\mathbf{r}_2,t_2)$.

This article deals with second-order statistical moments of fields; i.e., ensemble averages of such quadratic forms involving $\alpha(\mathbf{r},t)$, $\alpha(\mathbf{s},t)$, $\alpha(\mathbf{s},f)$, etc. Interest will center on the Fourier-equivalent two-point correlations $\Gamma_\alpha(\mathbf{r}_1,t_1,\mathbf{r}_2,t_2) = \langle \alpha^*(\mathbf{r}_1,t_1)\alpha(\mathbf{r}_2,t_2)\rangle$, $\Gamma_\alpha(\mathbf{r}_1,f_1,\mathbf{r}_2,f_2) = \langle \alpha^*(\mathbf{r}_1,f_1)\alpha(\mathbf{r}_2,f_2)\rangle$, or $\Gamma_\alpha(\mathbf{s}_1,f_1,\mathbf{s}_2,f_2) = \langle \alpha^*(\mathbf{s}_1,f_1)\alpha(\mathbf{s}_2,f_2)\rangle$.

The same basic symbol $\Gamma_\alpha(\cdots)$ is used for all of them, and only the arguments distinguish between the space/wave number and time/frequency domains. In dealing with statistically *stationary* fields, it is often convenient to transform to "mean and difference" coordinates; i.e., to re-express the dependence on a coordinate pair (ζ_i,ζ_j) in terms of a dependence on their mean $\bar{\zeta}_{ij} = \frac{1}{2}(\zeta_j + \zeta_i)$ and difference $\zeta_{ij} = \zeta_j - \zeta_i$. In this representation, the three moments above are $\Gamma_\alpha(\mathbf{r}_{12},t_{12},\bar{\mathbf{r}}_{12},\bar{t}_{12}) = \langle \alpha^*(\mathbf{r}_1,t_1)\alpha(\mathbf{r}_2,t_2)\rangle$, $\Gamma_\alpha(\mathbf{r}_{12},\bar{f}_{12},\bar{\mathbf{r}}_{12},f_{12}) = \langle \alpha^*(\mathbf{r}_1,f_1)\alpha(\mathbf{r}_2,f_2)\rangle$, and $\Gamma_\alpha(\bar{\mathbf{s}}_{12},f_{12},\mathbf{s}_{12},\bar{f}_{12}) = \langle \alpha^*(\mathbf{s}_1,f_1)\alpha(\mathbf{s}_2,f_2)\rangle$.

In the case of temporal stationarity, the standard terminology and notation are used; $C_\alpha(\mathbf{r}_1,\mathbf{r}_2,t_{12}) \equiv \Gamma_\alpha(\mathbf{r}_{12},t_{12},\bar{\mathbf{r}}_{12},\bar{t}_{12}) = C_\alpha^*(\mathbf{r}_2,\mathbf{r}_1,-t_{12})$ is the mutual coherence function, and its $\xrightarrow[t_{12} \quad f_{12}]{}$ transform $C_\alpha(\mathbf{r}_1,\mathbf{r}_2,\bar{f}_{12}) \equiv \Gamma_\alpha(\mathbf{r}_{12},\bar{f}_{12},\bar{\mathbf{r}}_{12},f_{12} \sim 0) = C_\alpha^*(\mathbf{r}_2,\mathbf{r}_1,\bar{f}_{12})$ is the cross-spectral density (CSD).

TABLE I. Elements of the problem considered as a linear system.

Input	$e(t)$	Emitted signal modulation (narrow band)
Internal time dependence	$h(\mathbf{r},t)$	Sea-surface elevation (slow relative to f_0^{-1})
Output	$A(\vec{r},t)$	Scattered field modulation (narrow band)

III. RANDOM FIELDS

We ultimately formulate the problem in terms of (i) a point source at $\vec{r} = \vec{r}_0$ that emits a narrow-band signal $e(t) \times E(-f_0 t)$; (ii) a sea surface $h(\mathbf{r},t)$ that evolves slowly on the acoustic time scale; and (iii) a narrow-band scattered field $A(\vec{r},t) \times E(-f_0 t)$. We are thus dealing with the linear time-dependent system specified in Table I.

A. Sea surface elevation

Even if the sea surface were completely arbitrary in form, it could still be Fourier analyzed according to

$$h(\mathbf{r},t) = \int_{\mathbb{R}^2} d^2\mathbf{s} \int_{-\infty}^{+\infty} df\, E(\mathbf{s}\cdot\mathbf{r} - ft) h(\mathbf{s},f), \tag{1}$$

with the restriction $h(\mathbf{s},f) = h^*(-\mathbf{s},-f)$ to guarantee that $h(\mathbf{r},t)$ remains real. Of course the traveling waves $E(\mathbf{s}\cdot\mathbf{r} - ft)$ that comprise the surface are not that arbitrary. They obey the appropriate dispersion relation for waves on the sea surface—a constraint of the form $|f| = F(s)$; i.e., $s = S(|f|)$ in terms of the inverse mapping, $S = F^{-1}$. (Recall that $s = |\mathbf{s}|$.) This shrinks the 3D integration region $\{\mathbf{s} \in \mathbb{R}^2, |f| < \infty\}$ of Eq. (1), to the 2D manifold $\{\mathbf{s} \in \mathbb{R}^2, |f| = F(s)\}$.

The sea surface can also be synthesized via

$$h(\mathbf{r},t) = \oint d\theta \int_0^\infty df\, \alpha(\theta,f) \cos\{\phi(\theta,f) + 2\pi[S(f)\mathbf{n}(\theta)\cdot\mathbf{r} - ft]\}, \tag{2}$$

in which the traveling wave component for each (θ,f) has an amplitude α and a phase ϕ, both real. [The notation $\mathbf{s} = s\mathbf{n}(\theta)$ is used, and $\oint d\theta(\cdots)$ denotes a 360° azimuth integration.] This incorporates the dispersion relation and is manifestly real valued. It also avoids double counting: since $f > 0$, the $(\theta,|f|)$ component is included, but its physical twin $(\theta+\pi,-|f|)$ is not.

Naturally, the two approaches are equivalent. The α and ϕ in Eq. (2), which are defined only for positive frequencies, can be extended to the negative spectrum through $\alpha(\theta,-f) = \alpha(\theta+\pi,f)$ and $\phi(\theta,-f) = -\phi(\theta+\pi,f)$. Then, in terms of the complex amplitude $u(\theta,f) \equiv \frac{1}{2}\alpha(\theta,f)e^{i\phi(\theta,f)}$ with symmetry $u(\theta,f) = u^*(\theta+\pi,-f)$, Eq. (2) is simply

$$h(\mathbf{r},t) = \int_{-\infty}^{+\infty} df \oint d\theta\, u(\theta,f) E[S(f)\mathbf{n}(\theta)\cdot\mathbf{r} - ft]. \tag{3}$$

Any physical sea surface can be synthesized in this way by integration over azimuth and frequency. The symmetry of u guarantees that $h(\mathbf{r},t)$ is real. Comparison of Eqs. (1) and (3) shows that

$$h(\mathbf{s},f)=s^{-1}\delta(s-S(f))u(\theta,f). \tag{4}$$

B. Scattered field

The acoustic pressure field satisfies the familiar differential boundary value problem

$$\begin{aligned} &WP(\vec{r},t)=Q(\vec{r},t) \quad z>h(\mathbf{r},t)\\ &P(\vec{r},t)=0 \quad z=h(\mathbf{r},t)\\ &P(\vec{r},t)\to 0 \quad |\vec{r}|\to\infty, \end{aligned} \tag{5}$$

in terms of the wave operator $W=-\{\rho(\vec{r})\partial/\partial\vec{r}\}\cdot\{\rho^{-1}(\vec{r})\partial/\partial\vec{r}\}+c^{-2}(\vec{r})\partial^2/\partial t^2$, in which ρ and c are the density and sound speed of the medium. The time dependence in $P(\vec{r},t)$ comes from the sea surface $h(\mathbf{r},t)$ and the source term $Q(\vec{r},t)$. We require that the latter remain fixed in position and have a uniform time dependence of the form $Q(\vec{r},t)=q(\vec{r})\times e(t)\exp(-i\omega_0 t)$. The signal emitted by this source consists of a carrier (frequency ω_0) with a modulation, $e(t)$, which is assumed to be slow on the time scale of the carrier. This gives the acoustic field the form $P(\vec{r},t)=A(\vec{r},t)\exp(-i\omega_0 t)$, in which $A(\vec{r},t)$ is a slow, complex-valued modulation. When the source is idealized as a point [by taking $q(\vec{r})=\delta(\vec{r}-\vec{r}_0)$] and the slowness of $A(\vec{r},t)$ is exploited [by neglecting $\partial^n A/\partial(\omega_0 t)^n$ relative to A], the result is

$$\begin{aligned} &\hat{W}A(\vec{r},t)=e(t)\,\delta(\vec{r}-\vec{r}_0) \quad z>h(\mathbf{r},t)\\ &A(\vec{r},t)=0 \quad z=h(\mathbf{r},t)\\ &A(\vec{r},t)\to 0 \quad |\vec{r}|\to\infty, \end{aligned} \tag{6}$$

in which $\hat{W}=-[\{\rho(\vec{r})\partial/\partial\vec{r}\}\cdot\{\rho^{-1}(\vec{r})\partial/\partial\vec{r}\}+k_0^2(\vec{r})]$ is the general Helmholtz operator with wave number $k_0(\vec{r})=\omega_0/c(\vec{r})$. The time dependence in A comes from both e and h. If one "freezes" the sea surface, i.e., disallows any time dependence in h, the ratio $A(\vec{r},t)/e(t)$ loses its time dependence too. (It is simply the space-frequency Green's function for the problem with the fixed boundary h at the frequency f_0, i.e., the response function of the now time-invariant environment at the carrier frequency.)

Slow time dependence can be retained in h by first converting Eq. (6) to integral form. By Green's theorem, $\hat{W}u(\vec{r})=\delta(\vec{r}-\vec{r}_0)$ within a volume V is equivalent to

$$u(\vec{r})=G(\vec{r}-\vec{r}_0)-\int_{\partial V} G_a(\vec{r},\vec{r}_a)u(\vec{r}_a)\,d^2\vec{r}_a, \tag{7}$$

in which $G(\vec{r},\vec{r}_a)$ is the Green's function for pressure-release conditions on the boundary ∂V, and $G_a(\vec{r},\vec{r}_a)$ is its outward normal derivative at the point $\vec{r}_a\in\partial V$. Equation (7) can be applied to $u=A/e$ by identifying V with the half-space $z>0$ and ∂V with the $z=0$ plane, h. The result is

$$A(\vec{r},t)=e(t)G(\vec{r},\vec{r}_0)-\int_h \mathrm{d}^2\vec{r}_a G_a(\vec{r},\vec{r}_a)A(\vec{r}_a,t), \tag{8}$$

in which G is the Green's function for pressure-release conditions on h, and the vertical surface derivative

$$G_a(\vec{r},\vec{r}_a)\triangleq -\left(\frac{\partial G(\vec{r},\vec{r}_a)}{\partial z_a}\right)_{z_a=0},$$

is the corresponding "boundary Green's function."[13]

A perturbative solution to Eq. (8) can be obtained through the small waveheight approximation,[14] which expresses the modulation in terms of the SWHA series $A=A^{(0)}+A^{(1)}+A^{(2)}+\cdots$ with $A^{(n)}=O(|h|^n)$. Applied to Eq. (8), this produces

$$A^{(0)}(\vec{r},t)=e(t)G(\vec{r},\vec{r}_0), \tag{9}$$

$$A^{(1)}(\vec{r},t)=-e(t)\int_h \mathrm{d}^2\mathbf{r}_a G_a(\vec{r},\vec{r}_a)h(\mathbf{r}_a,t)G_a(\vec{r}_a,\vec{r}_0), \tag{10}$$

$$\begin{aligned} A^{(2)}(\vec{r},t)=e(t)\int\!\!\int_h \mathrm{d}^2\mathbf{r}_a\mathrm{d}^2\mathbf{r}_b G_a(\vec{r},\vec{r}_a)h(\mathbf{r}_a,t)\\ \times G_{ab}(\vec{r}_a,\vec{r}_b)h(\mathbf{r}_b,t)G_b(\vec{r}_b,\vec{r}_0), \end{aligned} \tag{11}$$

etc. Equations (9)–(11) are the first three terms of the SWHA series for the complex modulation of a narrow-band signal that is scattered from a given time-dependent sea surface. This is emphatically *not* a new result [it is only a minor variation on Eqs. (30) and (31) of Ref. 6, for example]; however, it is a natural starting point for the following section's treatment in terms of the statistical moments of the random processes h, e, and A.

IV. STATISTICAL MOMENTS

A. Sea surface elevation

We assume that the random sea surface is stationary both in time and in horizontal coordinates. This implies that it consists of sinusoidal components that are uncorrelated with one another in both frequency and azimuth angle. Such an ensemble is sometimes called a "free wave field,"[15] and is customarily described in terms of its frequency spectrum and its azimuthal directionality.

1. Free-wave sea surface

In a stochastic description, the complex amplitude $u(\theta,f)$ is a random process. Since $\langle u\rangle=0$, the relevant statistic is the second moment

$$\Gamma_u(\theta_1,f_1,\theta_2,f_2)=\langle u^*(\theta_1,f_1)u(\theta_2,f_2)\rangle. \tag{12}$$

From the traveling-wave surface representation, Eq. (3), we have

$$\begin{aligned} &\Gamma_h(\mathbf{r}_1,t_1,\mathbf{r}_2,t_2)\\ &=\oint\oint \mathrm{d}\theta_1\,\mathrm{d}\theta_2\int\!\!\int \mathrm{d}f_1\,\mathrm{d}f_2\langle u^*(\theta_1,f_1)u(\theta_2,f_2)\rangle\\ &\quad\times E[-S(f_1)\mathbf{n}(\theta_1)\cdot\mathbf{r}_1+S(f_2)\mathbf{n}(\theta_2)\cdot\mathbf{r}_2+f_1t_1-f_2t_2]\\ &=\int\!\!\int \mathrm{d}^2\mathbf{s}_1\mathrm{d}^2\mathbf{s}_2 s_1^{-1}\delta(s_1-S(f_1))s_2^{-1}\delta(s_2-S(f_2))\\ &\quad\times\Gamma_u(\theta_1,f_1,\theta_2,f_2)E[(\overline{\mathbf{s}}_{12}\cdot\mathbf{r}_{12}+\mathbf{s}_{12}\cdot\overline{\mathbf{r}}_{12})\\ &\quad-(\overline{f}_{12}t_{12}+f_{12}\overline{t}_{12})], \end{aligned} \tag{13}$$

in the second form of which we have used the polar representation $\mathbf{s}_j = s_j \mathbf{n}(\theta_j)$ and introduced mean and difference coordinates. We have supposed that $h(\mathbf{r},t)$ is statistically stationary in t, which is commonly a valid approximation. As a consequence, $\Gamma_h(\mathbf{r}_{12},t_{12},\bar{\mathbf{r}}_{12},\bar{t}_{12})$ must be independent of $\bar{t}_{12}$ and therefore the factor $E(-f_{12}\bar{t}_{12})$ cannot actually introduce any $\bar{t}_{12}$ dependence, no matter what value f_{12} may take on. In other words, a factor $\delta(f_{12})$ must be present. We are also supposing that $h(\mathbf{r},t)$ is stationary in $\mathbf{r}$. By similar reasoning, there must also be a factor $\delta(\mathbf{s}_{12}) = s_1^{-1}\delta(\theta_1 - \theta_2)\delta(s_1 - s_2)$ present. Thus, the traveling-wave components of the sea surface are uncorrelated in both frequency and azimuth; i.e., they constitute a free wave field. Consequently it is not difficult to show [using the polar form $\bar{\mathbf{s}}_{12} = \bar{s}_{12}\mathbf{n}(\bar{\theta}_{12})$] that

$$C_h(\bar{\mathbf{s}}_{12},\mathbf{s}_{12}\sim\mathbf{0},\bar{f}_{12}) = \frac{1}{\bar{s}_{12}}\,\delta(\bar{s}_{12} - S(\bar{f}_{12}))C_u(\bar{\theta}_{12},\theta_{12}\sim 0,\bar{f}_{12}), \tag{14}$$

the form and content of which are a direct inheritance from Eq. (4).

2. Surface frequency spectrum and azimuthal directionality

It is "customary"[16] to express the angle-frequency moment in Eq. (14) in factored form

$$C_u(\bar{\theta}_{12},\theta_{12}\sim 0,\bar{f}_{12}) = \Phi(\bar{\omega}_{12})\begin{cases} H(\bar{\theta}_{12}) & \text{for } \bar{\omega}_{12} > 0 \\ H(\bar{\theta}_{12}+\pi) & \text{for } \bar{\omega}_{12} < 0 \end{cases}, \tag{15}$$

in terms of the sea-surface frequency spectrum $\Phi(\omega)$ and azimuthal directionality $H(\theta)$, both of which are real and positive. In many treatments, a frozen sea surface is assumed, and thus only the positive-frequency part of Eq. (15) appears [e.g., Ref. 16, Eq. (3.22)]. That will not be adequate here because the aim is to deal with Doppler effects.

B. Scattered field

The second moment of the scattered field modulation in the space/time domain is

$$\begin{aligned}\langle A^*(\vec{r}_1,t_1)A(\vec{r}_2,t_2)\rangle &= \Gamma_A(\vec{r}_1,t_1,\vec{r}_2,t_2) \\ &\equiv \Gamma_A(\mathbf{r}_{01},t_{12},\mathbf{r}_{02},\bar{t}_{12};z_0,z_1,z_2).\end{aligned} \tag{16}$$

In the final form, the dependence on source position $\vec{r}_0 = (\mathbf{r}_0,z_0)$ is explicit, mean and difference time coordinates (t_{12},t_{12}) are used, and environmental range invariance is exploited to express the range dependence in terms of $\mathbf{r}_{01}$ and $\mathbf{r}_{02}$ rather than $\mathbf{r}_0$, $\mathbf{r}_1$, and $\mathbf{r}_2$ separately. The $\langle\cdots\rangle$ brackets denote averaging over sea surfaces (and also over source modulations in the case of a stochastic source; i.e., when $e(t)$ is a random process).

The SWHA series for A leads directly to $\Gamma_A = \Gamma_A^{(0)} + \Gamma_A^{(1)} + \Gamma_A^{(2)} + \cdots$, an SWHA series for the second moment. The form is the same in the wave number and frequency domains. The space/frequency version in particular is obtained by the transformation

$$\begin{aligned}&\Gamma_A(\mathbf{r}_{01},t_{12},\mathbf{r}_{02},\bar{t}_{12};\cdots) \\ &\quad \overset{t_{12}\ \bar{f}_{12}}{\longrightarrow}\ \overset{\bar{t}_{12}\ f_{12}}{\longrightarrow} \Gamma_A(\mathbf{r}_{01},\bar{f}_{12},\mathbf{r}_{02},f_{12};\cdots),\end{aligned} \tag{17}$$

which, in the time-stationary case, reduces to the CSD

$$C_A(\mathbf{r}_{01},\mathbf{r}_{02},t_{12};\cdots) \overset{t_{12}\ \bar{f}_{12}}{\longrightarrow} C_A(\mathbf{r}_{01},\mathbf{r}_{02},\bar{f}_{12};\cdots). \tag{18}$$

For simplicity, the depths in both of these expressions have been abbreviated by "$\cdots$."

We are interested in Doppler effects in $C_A = C_A^{(0)} + C_A^{(1)} + C_A^{(2)} + \cdots$, the SWHA series for the CSD. The first two terms are of no interest since $C_A^{(0)}$ is the main zero-Doppler contribution and $C_A^{(1)}$ vanishes. The next term, $C_A^{(2)}$, contains contributions of two types: $\langle A^{(2)^*}A^{(0)}\rangle$ and $\langle A^{(1)^*}A^{(1)}\rangle$. The first is merely a small base-band (i.e., zero-Doppler) correction, while the second provides the Doppler sidebands. We will hereafter drop the superscript $^{(2)}$, with the understanding that we are always concerned solely with this second-order Doppler contribution.

The Doppler contribution to the second-order space/time moment is

$$\begin{aligned}\Gamma_A(\mathbf{r}_{01},t_1,\mathbf{r}_{02},t_2;\cdots) &= \Gamma_e(t_1,t_2) \\ &\times \int\int \mathrm{d}\mathbf{r}_a\,\mathrm{d}\mathbf{r}_b G_a^*(\mathbf{r}_{a1};z_1,z_a)G_a^*(\mathbf{r}_{0a};z_a,z_0) \\ &\times G_b(\mathbf{r}_{b2};z_2,z_b)G_b(\mathbf{r}_{0b};z_b,z_0)\Gamma_h(\mathbf{r}_{0a},t_1,\mathbf{r}_{0b},t_2).\end{aligned} \tag{19}$$

This can be written as $\Gamma_A(\mathbf{r}_{01},t_{12},\mathbf{r}_{02},\bar{t}_{12};\cdots) = \Gamma_e(t_{12},\bar{t}_{12})\mathcal{L}\Gamma_h(\mathbf{r}_{0a},t_{12},\mathbf{r}_{0b},\bar{t}_{12})$ in mean and difference time coordinates, with $\mathcal{L}$ denoting the integral operator in Eq. (19). The stationarity of the source implies that $\Gamma_e(t_{12},\bar{t}_{12}) = C_e(t_{12})$; the stationarity of the sea surface indicates that $\Gamma_h(\mathbf{r}_{0a},t_{12},\mathbf{r}_{0b},\bar{t}_{12};\cdots) = C_h(\mathbf{r}_{0a},\mathbf{r}_{0b},t_{12};\cdots)$; and together they imply that $A^{(1)}(\vec{r},t)$ is stationary in time and that $\Gamma_A(\mathbf{r}_{01},t_{12},\mathbf{r}_{02},\bar{t}_{12};\cdots) = C_A(\mathbf{r}_{01},\mathbf{r}_{02},\bar{t}_{12};\cdots)$.

In the frequency domain, therefore

$$C_A(\mathbf{r}_{01},\mathbf{r}_{02},f;\cdots) = C_e(f) \underset{f}{\star} \hat{C}_A(\mathbf{r}_{01},\mathbf{r}_{02},f;\cdots), \tag{20}$$

in which the source PSD is convolved with the scattered "reference" CSD (the CSD produced by a unit-strength, zero-bandwidth source)

$$\begin{aligned}\hat{C}_A(\mathbf{r}_{01},\mathbf{r}_{02},f;z_0,z_1,z_2) &= \int\int \mathrm{d}^2\mathbf{r}_a\,\mathrm{d}^2\mathbf{r}_b G_a^*(\mathbf{r}_{a1};z_1,z_a)\{G_a^*(\mathbf{r}_{0a};z_a,z_0) \\ &\times C_h(\mathbf{r}_{ab},\bar{\mathbf{r}}_{ab},f)G_b(\mathbf{r}_{0b};z_b,z_0)\}G_b(\mathbf{r}_{b2};z_2,z_b).\end{aligned} \tag{21}$$

For a *passive* problem, with the sea surface acting as the *primary* acoustic source, the curly brackets in Eq. (21) would contain nothing but the boundary-source CSD [Eq. (3.19) of Ref. 17]. Thus, Eq. (21) would reduce to a form of the van Cittert–Zernike theorem from the classical theory of partially coherent wave fields (Sec. 10.4.2 of Ref. 18), which expresses the CSD of a narrow-band field within a volume in terms of its values on the bounding surface. This can be a

convenient starting point for passive problems [e.g., in the Kuperman and Ingenito treatment of the surface-generated ambient noise field, Eq. (7) of Ref. 19]. In the *active* problem, the sea surface is a scatterer; i.e., a *secondary* source. In this case, rather than that simple boundary value, the curly brackets contain (a) two Green's functions that propagate the signal from the primary source to a pair of surface points; and (b) a waveheight CSD that embodies the sea-surface correlation between those surface points. Equation (21) is, in that sense, a generalization of the van Cittert–Zernike theorem to active scattering from a time-dependent surface.

When the angle-frequency moment has the factored form Eq. (15), Doppler sidebands emerge (after some fairly tedious 2D Fourier transform manipulations[20]) as

$$\hat{C}_A(\mathbf{r}_{01},\mathbf{r}_{02},\pm|f|;z_0,z_1,z_2) = \Phi(|\omega|)\oint \mathrm{d}\theta\, H(\theta)[\mathrm{e}^{\mathrm{i}\mathbf{r}\cdot\mathbf{K}/2} I^*(\mathbf{K},\mathbf{r}_{01};z_0,z_1) \times I(\mathbf{K},\mathbf{r}_{02};z_0,z_2)]_{\mathbf{K}=\pm K\mathbf{n}(\theta)}, \tag{22}$$

in terms of the surface wave dispersion relation $|\omega|=\Omega(K)$ for $K\geqslant 0$, and in which

$$I(\mathbf{K},\mathbf{r};z_0,z) = (2\pi)^{-2}\int \mathrm{d}^2\mathbf{k} e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} G_\alpha(\mathbf{k}-\tfrac{1}{2}\mathbf{K};z_\alpha,z_0) \times G_\alpha(\mathbf{k}+\tfrac{1}{2}\mathbf{K};z,z_\alpha). \tag{23}$$

Since the wave vector is restricted to $\mathbf{K}=\pm K\mathbf{n}(\theta)$, each surface wave number contributes explicitly to both the upper $(+)$ and lower $(-)$ sidebands. For each K, Eq. (22) is to be evaluated by integrating over azimuth and then assigning the result to the proper Doppler-shifted frequency, $\pm|f|$. Although formulations of this general type have been used with *frozen* sea surfaces,[12] this is, as far as the author is aware, the first to explicitly separate out the Doppler sidebands in the scattering from a moving surface.

The only real difficulty in applying the above formulation is the propagation integral I, Eq. (23). If an analytic form is available for I, Eq. (22) requires only a single numerical integration (over azimuth). Otherwise, all of the integration must be done numerically—a significantly more difficult proposition. Fortunately, some environments are simple enough that, with effort, Eq. (23) *can* be evaluated analytically. The following section treats the simplest such environment, a uniform ocean.

C. Homogeneous ocean case

The Green's function for a homogeneous ocean with sound speed c_0 has the form

$$G(\mathbf{k};z,z') = \underbrace{\frac{\exp(\mathrm{i}\kappa(k_0,k)|z'-z|)}{2\mathrm{i}\kappa(k_0,k)}}_{\text{source}} - \underbrace{\frac{\exp(\mathrm{i}\kappa(k_0,k)|z'+z|)}{2\mathrm{i}\kappa(k_0,k)}}_{\text{image}}, \tag{24}$$

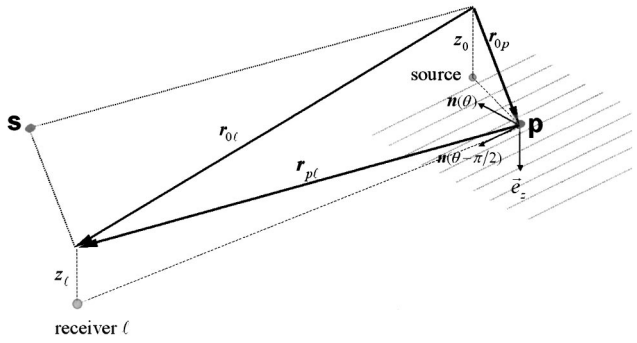


FIG. 1. Bragg scattering at surface point **p** by a surface wave having wave vector $\mathbf{K}=\pm K\mathbf{n}(\theta)$. The wave crests (parallel lines) are spaced $\Lambda=2\pi/K$ apart. The symmetric surface point **s** is also shown.

in which the acoustic wave number and its horizontal and vertical components are $k_0=|\vec{k}|=\omega_0/c_0$, $k=|\mathbf{k}|$, and $\kappa(k_0,k)=\sqrt{k_0^2-k^2}$, respectively. The source and surface image terms correspond to wave vectors $\vec{k}=\mathbf{k}+\kappa(k_0,k)\vec{e}_z$ and $\vec{k}=\mathbf{k}-\kappa(k_0,k)\vec{e}_z$. Surface derivatives are simply $G_a(\mathbf{k};z_b,z_a)=\exp(\mathrm{i}\kappa(k_0,k)z_b)$ and $G_b(\mathbf{k};z_b,z_a)=\exp(\mathrm{i}\kappa(k_0,k)z_a)$. This leaves the propagation integral in the form

$$I(\mathbf{K},\mathbf{r}_{0\ell};z_0,z_\ell) = (2\pi)^{-2}\int \mathrm{d}^2\mathbf{k}\exp(\mathrm{i}\Psi(\mathbf{k},\mathbf{K},k_0;\mathbf{r}_{0\ell},z_0,z_\ell)) \quad \text{for } \ell=1,2, \tag{25}$$

in which

$$\Psi(\mathbf{k},\mathbf{K},k_0;\mathbf{r}_{0\ell},z_0,z_\ell) = \mathbf{k}\cdot\mathbf{r}_{0\ell} + \kappa(k_0,|\mathbf{k}-\tfrac{1}{2}\mathbf{K}|) + \kappa(k_0,|\mathbf{k}+\tfrac{1}{2}\mathbf{K}|), \tag{26}$$

for any surface wave vector **K**. Assuming Eq. (26) is real, Eq. (25) can be evaluated by applying stationary-phase methods to the phase, Ψ.

The stationarity condition, $\partial\Psi(\mathbf{k},\cdots)/\partial\mathbf{k}=\mathbf{0}$, amounts to

$$\mathbf{r}_{0\ell} = z_0\frac{\mathbf{k}-\frac{1}{2}\mathbf{K}}{\kappa(k_0-|\mathbf{k}-\frac{1}{2}\mathbf{K}|)} + z_\ell\frac{\mathbf{k}+\frac{1}{2}\mathbf{K}}{\kappa(k_0-|\mathbf{k}+\frac{1}{2}\mathbf{K}|)}. \tag{27}$$

Any **k** that solves Eq. (27) for the prevailing **K** defines a sea surface point **p** via $\mathbf{r}_{0\ell}=\mathbf{r}_{0p}+\mathbf{r}_{p\ell}$ (Fig. 1), with

$$\frac{\mathbf{r}_{0p}}{z_0} = \frac{\mathbf{k}-\frac{1}{2}\mathbf{K}}{\kappa(k_0,|\mathbf{k}-\frac{1}{2}\mathbf{K}|)} \quad \text{and} \quad \frac{\mathbf{r}_{p\ell}}{z_\ell} = \frac{\mathbf{k}+\frac{1}{2}\mathbf{K}}{\kappa(k_0,|\mathbf{k}+\frac{1}{2}\mathbf{K}|)}. \tag{28}$$

These correspond to propagation to and from **p** with 3D wave vectors

$$\underbrace{\vec{k}}_{\text{to}} = (\mathbf{k} - \tfrac{1}{2}\mathbf{K}) - \kappa(k_0, |\mathbf{k} - \tfrac{1}{2}\mathbf{K}|)\vec{e}_z$$

and

$$\underbrace{\vec{q}}_{\text{from}} = (\mathbf{k} + \tfrac{1}{2}\mathbf{K}) - \kappa(k_0, |\mathbf{k} + \tfrac{1}{2}\mathbf{K}|)\vec{e}_z . \tag{29}$$

Scattering at **p** produces a change

$$\vec{Q} \equiv \vec{k} - \vec{q} = \mathbf{K} + [\kappa(k_0, |\mathbf{k} + \tfrac{1}{2}\mathbf{K}|) + \kappa(k_0, |\mathbf{k} - \tfrac{1}{2}\mathbf{K}|)]\vec{e}_z , \tag{30}$$

in the wave vector. Energy conservation requires

$$|\vec{k}| = |\vec{q}| = k_0 , \tag{31}$$

and, from Eq. (30) with $\mathbf{K} = \pm K\mathbf{n}(\theta)$, one has

$$\vec{Q} \cdot \mathbf{n}(\theta \pm \tfrac{1}{2}\pi) = 0 \quad \text{and} \quad |\vec{Q} \cdot \mathbf{n}(\theta)| = K. \tag{32}$$

Equations (31) and (32) constitute the condition for the signal to travel from the source to receiver ℓ via first-order Bragg scattering at point **p** involving the surface-wave component with wave vector **K**. If one swaps the values of $\mathbf{r}_{0p}$ and $\mathbf{r}_{p\ell}$, this scattering occurs instead at the point **s** with $\mathbf{r}_{0s} = \mathbf{r}_{p\ell}$ and $\mathbf{r}_{s\ell} = \mathbf{r}_{0p}$. Thus, the stationary-phase approximation corresponds to Bragg scattering at points **p** and **s** which lie on the port and starboard sides of the source-to-receiver "heading" $\mathbf{r}_{0\ell}$ (Fig. 1).

Since Ψ, Eq. (26), is to remain real, the **k** at which $\partial\Psi/\partial\mathbf{k} = \mathbf{0}$ must be sought within the region where both $\kappa(k_0, |\mathbf{k} + \frac{1}{2}\mathbf{K}|)$ and $\kappa(k_0, |\mathbf{k} - \frac{1}{2}\mathbf{K}|)$ are real valued; i.e., in the overlap of the two **k**-space disks $|\mathbf{k} + \frac{1}{2}\mathbf{K}| < k_0$ and $|\mathbf{k} - \frac{1}{2}\mathbf{K}| < k_0$. In Ref. 20, the stationary-phase evaluation is done analytically with the source and receivers at arbitrary horizontal positions but with their depths small relative to the source/receiver ranges. The analysis exploits bipolar coordinates, a coordinate system that exactly conforms to this overlapping-disks geometry. Full details (too lengthy for inclusion here) are to be found there. The outcome is that Eq. (22) reduces to

$$\hat{C}_A(\mathbf{r}_{01}, \mathbf{r}_{02}, \pm|f|; z_0, z_1, z_2) = A e^{ik_0(r_{02} - r_{01})} \int_{F(K/2k_0)} d\theta \left\{ BH\left(\theta - \frac{\pi}{2} \pm \frac{\pi}{2}\right) + De^{-iK\rho\cos(\theta - \varphi)} H\left(\theta + \frac{\pi}{2} \pm \frac{\pi}{2}\right) \right\}, \tag{33}$$

in which

$$A = \frac{\Phi(|\omega|)}{r_{01} r_{02}} \left(\frac{k_0}{4\pi}\right)^2, \quad B = \frac{z_1 z_2}{r_{01} r_{02}}, \quad D = \frac{z_0^2}{r_{01} r_{02}}, \tag{34}$$

$\mathbf{r}_{12} = \rho\mathbf{n}(\varphi)$ is the horizontal displacement of the second receiver relative to the first, and $\theta \in F(K/2k_0)$ is the azimuthal sector within which the surface wave $\mathbf{K} = K\mathbf{n}(\theta)$ is able to Bragg scatter the signal from the source to both receivers. The sea surface is specified in terms of (i) its dispersion relation $|\omega| = \Omega(K)$; (ii) its power spectrum $\Phi(|\omega|)$; and (iii) its azimuthal directionality $H(\theta)$.

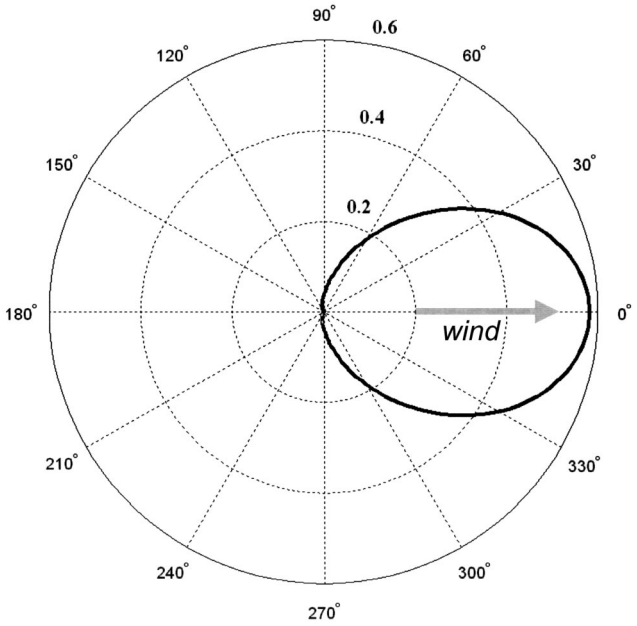


FIG. 2. Azimuthal directionality, Eq. (35), with $\mu = 4$.

V. EXAMPLE

This section provides an example of the computation of CSD sidebands using Eq. (33). The dispersion relation and surface power spectrum for the example are standard forms: $\Omega(K) = \sqrt{gK}$ corresponds to gravity waves in deep water, and

$$\Phi(\omega) = \frac{\alpha g^2}{\omega^5} \exp\left[-\frac{5}{4}\left(\frac{\omega_d}{\omega}\right)^4 \right]$$

is the Pierson–Moskowitz distribution,[21] an archetype for a fully developed wind-driven sea. The dominant-wave frequency ω_d depends inversely on the wind speed, U. The azimuthal directionality is taken to be

$$H(\theta) \propto |\cos(\tfrac{1}{2}\theta)|^{2\mu}, \tag{35}$$

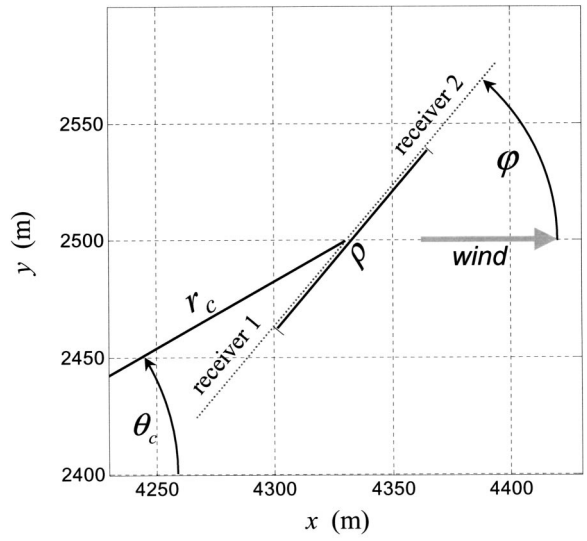


FIG. 3. Plan view of receiver geometries used in the example (see the text). The receiver separations are $\rho = 1,2,...,200$ m. ($\rho = 100$ m is illustrated.)

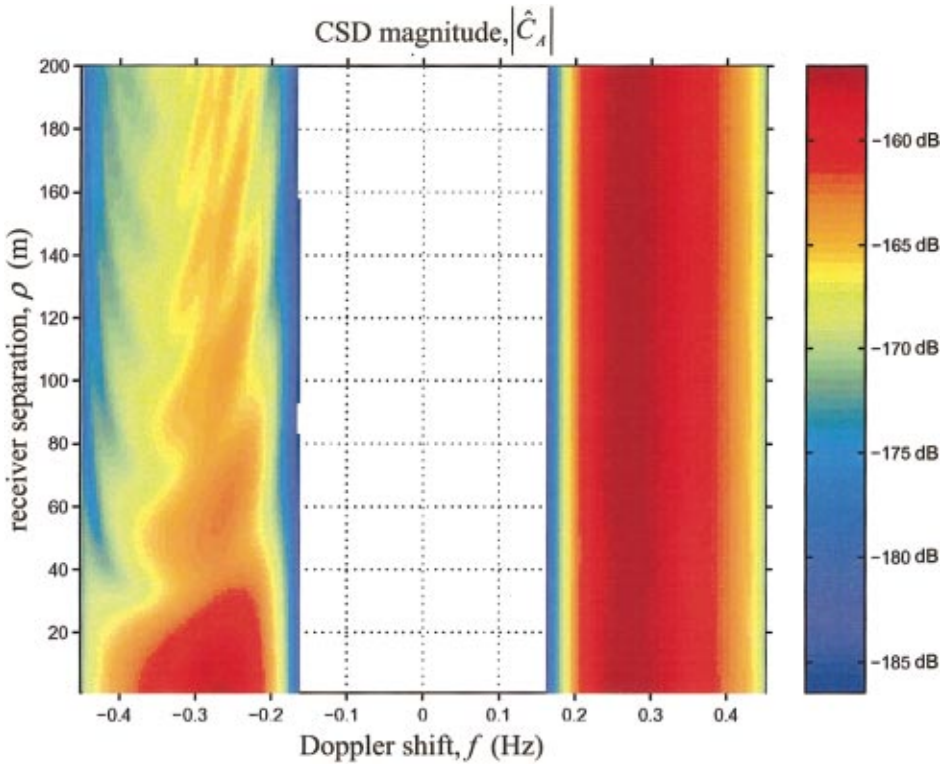


FIG. 4. CSD magnitude versus Doppler shift and receiver separation.

an empirical form devised by Longuet–Higgins and coworkers to fit at-sea data (with $\mu=4$, a typical value).[22,23] The normalization is $\oint H(\theta)\mathrm{d}\theta=1$. Unlike such forms as[12] $H(\theta)\propto\cos^2\theta$, Eq. (35) is *not* upwind/downwind symmetric (except of course in the isotropic limit, $\mu=0$). Rather, it is concentrated in the downwind direction ($\theta=0$, Fig. 2).

The computation is done at $f_0=100\,\mathrm{Hz}$ with $U=5$ m/s, $z_0=100$ m, and $z_1=z_2=90$ m. It covers a large set of receiver placements that are laid out relative to a common horizontal reference point $\mathbf{r}_c=(r_c,\theta_c)$ in the following way. The first $\mathbf{r}_1$, $\mathbf{r}_2$ pair closely brackets $\mathbf{r}_c$. Successive $\mathbf{r}_1$, $\mathbf{r}_2$ pairs separate farther along a horizontal line, $\mathbf{r}_{12}=\rho\mathbf{n}(\varphi)$, keeping $\mathbf{r}_c$ at their midpoint. The result is a series of receiver pair placements with a common value of φ and with steadily increasing ρ (Fig. 3). For these calculations, $r_c=5$ km, $\theta_c=30°$, $\varphi=50°$, and $\rho=1,2,...,200$ m.

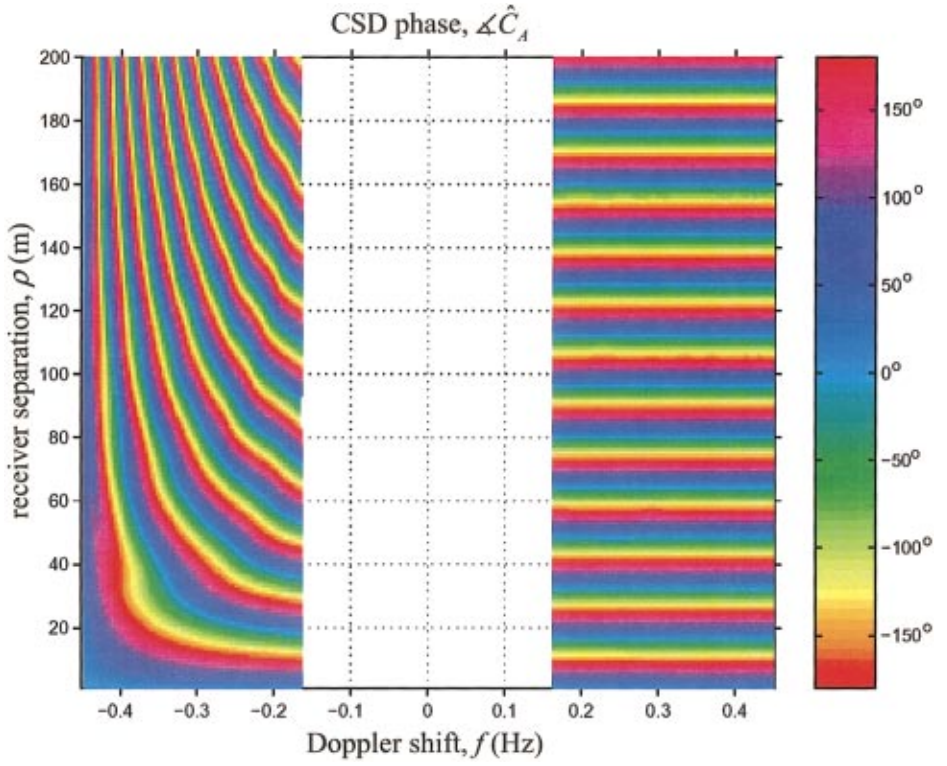


FIG. 5. CSD phase versus Doppler shift and receiver separation.

The complex CSD, $\hat{C}_A = |\hat{C}_A| \times \exp(i \measuredangle \hat{C}_A)$, is calculated using Eq. (33). Its magnitude $|\hat{C}_A|$ is plotted in Fig. 4 as a function of Doppler shift f and receiver separation ρ. The complete CSD would, of course, have a zero-Doppler component. The region around $f=0$ in this figure is blank because we are modeling only the sidebands. Absolute decibel levels are much lower than would be the case with a surface scattering strength because the propagation losses due to geometric spreading have not been removed. For all ρ, the upper sideband has a pronounced peak at $f \approx 0.27$ Hz, which corresponds to the dominant-wave frequency at this wind speed. The equivalent peak in the lower sideband is more variable and weaker by as much as 10 dB. This difference is ultimately traceable to the fact that the receivers lie downwind of the source in a fairly directional sea. The phase $\measuredangle \hat{C}_A$ appears in Fig. 5 in a similar format. In the upper sideband, each ρ yields the same phase at all Doppler shifts: the phase simply loops through about $12\frac{1}{2}$ regular 360° cycles as ρ increases from 1 to 200 m. This can be understood by reference to Fig. 3. The line containing the receivers makes an angle of $|\varphi - \theta_c| = 20°$ with the radial to the source. Since the acoustic wavelength is 15 m, direct line-of-sight propagation would predict a value of $\arccos(12.5 \times 15/200) = 20.36°$ for this angle, which agrees very well considering that it neglects the deviation produced by scattering.

VI. SUMMARY

A method has been presented for computing the Doppler sidebands in the cross-spectral density (CSD) of the acoustic field that is scattered by a general random, time-dependent sea surface. The general bistatic problem was posed with a narrow-band source and a pair of receivers, all deployed beneath a dynamic surface wave field. A perturbative method (the small waveheight approximation) was used to express the sideband contribution to the reverberation CSD as a double integral over the mean sea surface. Temporal and spatial stationarity reduced this to an extension [Eq. (21)] of the van Cittert–Zernike theorem that applies to active scattering from moving sea surfaces. The customary angle/frequency factorization ansatz [Eq. (15)] was invoked to produce a novel expression [Eq. (22)] for the up/down sidebands in which the sea-surface properties are embodied in three functions: the dispersion relation, the power spectrum, and the azimuthal directionality. This is the first explicit prediction of the Doppler sideband structure in the CSD of the field scattered from a realistic moving sea surface. For shallow source/receiver deployments in a uniform ocean, the fundamental 2D wave number integral [Eq. (23)] was evaluated by the stationary phase method, thus reducing the outcome to an easily implemented expression involving only a single integration (in azimuth). This was illustrated in a computational example that modeled the amplitude and phase dependence of both CSD sidebands on frequency and receiver separation for scenario with a pair of receivers deployed downwind from the source under a fully developed, wind-driven sea surface. Although the work assumes an ideal source (with unit strength and zero bandwidth), coverage can be extended to any real narrow-band source through simple frequency convolution with the source power spectrum [Eq. (20)]. Further developments are under consideration, primarily replacing the omnidirectional source by one with a beam pattern and introducing depth dependence into the ocean.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

[1] P. M. Ogden and F. T. Erskine, "Surface scattering measurements using broadband explosive charges in the Critical Sea Test experiments," J. Acoust. Soc. Am. **95**, 746–761 (1994).
[2] R. C. Gauss, R. F. Gragg, R. W. Nero, D. Wurmser, and J. M. Fialkowski, "Broadband models for predicting bistatic bottom, surface, and volume scattering strengths," Naval Research Laboratory Report 10042 (September, 2002).
[3] W. S. Burdic, *Underwater Acoustic System Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1984), Chaps. 10, 11.
[4] R. F. Gragg and B. H. Pasewark, "Stabilized high-resolution beamforming with horizontal arrays: Two experimental trials in shallow water," J. Acoust. Soc. Am. **89**, 682–690 (1991).
[5] C. S. Hayek, I. W. Schuman, J. H. Sweeney, and C. A. Boyles, "Azimuthal dependence of Bragg scattering from the ocean surface," J. Acoust. Soc. Am. **105**, 2129–2141 (1999).
[6] F. M. Labianca and E. Y. Harper, "Connection between various small-waveheight solutions of the problem of scattering from the ocean surface," J. Acoust. Soc. Am. **57**, 1044–1050 (1975).
[7] E. Y. Harper and F. M. Labianca, "Scattering of sound from a point source by a rough surface progressing over an isovelocity ocean," J. Acoust. Soc. Am. **58**, 349–364 (1975).
[8] F. M. Labianca and E. Y. Harper, "Asymptotic theory of scattering by a rough surface progressing over an inhomogeneous ocean," J. Acoust. Soc. Am. **59**, 799–812 (1976).
[9] F. M. Labianca and E. Y. Harper, "Sideband structure of sound from a harmonic point source scattered by a rough surface moving over an upward-refracting ocean," J. Acoust. Soc. Am. **61**, 378–389 (1977).
[10] F. M. Labianca and E. Y. Harper, "Connection between various small-waveheight solutions of the problem of scattering from the ocean surface," J. Acoust. Soc. Am. **62**, 1144–1157 (1977).
[11] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures* (Wiley-Interscience, New York, 1971), Chap. 5.
[12] P. W. Goalwin, "A stationary phase approach to the calculation of the correlation of the acoustic field scattered from a two-dimensional random sea surface," J. Acoust. Soc. Am. **93**, 214–223 (1993).
[13] P. M. Morse and M. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 807.
[14] J. W. Strutt Lord Rayleigh, "On the dynamical theory of gratings," Proc. R. Soc. London, Ser. A **79**, 399–416 (1907).
[15] S. M. Baxter and C. L. Morfey, *Angular Distribution Analysis in Acoustics* (Springer, Berlin, 1986), Chap. 2.
[16] M. A. Donelan, J. Hamilton, and W. H. Hui, "Directional spectra of wind-generated waves," Philos. Trans. R. Soc. London, Ser. A **315**, 509–562 (1985).
[17] J. Peřina, *Coherence of Light*, 2nd ed. (Reidel, Dordrecht, Holland, 1985).
[18] M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon, Oxford, 1980).
[19] W. A. Kuperman and F. Ingenito, "Spatial correlation of surface generated noise in a stratified ocean," J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[20] R. F. Gragg, "Space-frequency correlations in multistatic acoustic reverberation due to a wind-driven sea surface: Theoretical results at low frequency," Naval Research Laboratory Report 9923, November, 1999.
[21] W. J. Pierson and L. Moskowitz, "A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii," J. Geophys. Res. **69**, 5181–5190 (1965). (U is the wind speed at 19.5 m, g is the gravity constant, $\alpha = 0.0081$ is the Phillips constant, $\beta = 0.74$, and $\omega_d = (4\beta/5)^{1/4} g/U$.)
[22] M. S. Longuet-Higgins, D. E. Cartwright, and D. N. Smith, "Observations of the directional spectrum of sea waves using the motions of a floating buoy," *Proceedings of the Conference on Ocean Wave Spectra* (Prentice-Hall, Englewood Cliffs, NJ, 1963), pp. 111–132.
[23] H. Mitsuyasu, F. Tasai, T. Suhara, S. Mizuno, M. Ohkusu, T. Honda, and K. Rikiishi, "Observations of the directional spectrum of ocean waves using a cloverleaf buoy," J. Phys. Oceanogr. **5**, 750–760 (1975).

Broadband performance of a moving time reversing array[a)]

Karim G. Sabra and David R. Dowling[b)]
Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2121

(Received 13 September 2002; revised 25 June 2003; accepted 8 July 2003)

Acoustic time reversal exploits reciprocity between sources and receivers to generate backward propagating waves that automatically focus at their point of origin. In underwater acoustics, an array of transducers that can both transmit and receive, referred to as a time reversing array (TRA) or time reversal mirror (TRM), generates the back-propagating waves. Such arrays have been shown to spatially and temporally focus sound in unknown complicated multipath environments, and are therefore of interest for active sonar and underwater communication applications. Although stationary vertical linear TRAs have been favored in prior studies, practical applications of acoustic time reversal in underwater environments are likely to involve towed, tilted, horizontal, or bottom-mounted arrays. In particular, array motion introduces Doppler effects and eliminates source–receiver reciprocity, two factors that potentially impact the automatic focusing capability of TRAs. This paper presents the results from a theoretical and computational investigation into how array motion and orientation influence TRA retrofocusing in the shallow ocean. Here, the TRA tow speed is assumed constant, and the array is assumed to be straight and linear (vertical, horizontal, or tilted). And, for simplicity, the TRA is assumed to respond to a stationary point source emitting a broadband pulse. When a TRA moves, the retrofocus is predicted to shift in the direction of array motion due to the translation of the array between its reception and broadcast times. In addition, the performance of a towed horizontal TRA is predicted to degrade more rapidly with towing speed than that of an equivalent (but clearly idealized) towed vertical array because of range-dependent Doppler phase differences that do not influence the vertical array. However, short tilted arrays may approach vertical array performance and appear to be a potentially versatile compromise for implementing TRA concepts in active sonar or underwater communication systems. [DOI: 10.1121/1.1604124]

PACS numbers: 43.30.Es, 43.30.Vh, 43.30.Yj [WLS]

I. INTRODUCTION

Active acoustic time reversal is a technique for focusing sounds recorded in complex unknown environments back to their remote point(s) of origin. It is based on the invariance of the lossless wave equation to changing the sign of the time variable. Active acoustic time reversal may be accomplished by recording sounds with an array of transducers—a time-reversing array (TRA) or time-reversal mirror (TRM)—and then replaying the recorded sounds from the same array to produce back-propagating waves that converge at the location(s) of the remote sound source(s). TRAs may be of nearly any size or shape and can operate in any frequency range without knowledge of the source location(s), transducer locations, or environmental characteristics. TRAs perform well in the absence of acoustic absorption losses and temporal changes in the environment when the array possesses sufficient aperture. Future active sonar and underwater communication systems suitable for use in unknown shallow ocean waters may be developed from the automatic spatial and temporal focusing properties of TRAs.

Nearly all analysis and measurements of TRA performance in underwater environments have involved stationary vertical linear arrays. Recent studies of the time-reversal process include its experimental implementation in the ocean,[1] quantification of the spatial and temporal properties of the retrofocused field,[2,3] quantification of its sensitivity to ocean dynamics,[4] and ambient noise.[5,6] However, practical implementation of TRA technology for commercial or military sonar systems may require use of more nearly horizontal arrays that are bottom mounted or towed through the ocean. Thus, the combined effects of array motion and orientation on the performance of the time-reversal process are important for applications yet are largely unquantified. Array orientation is known to play a role in stationary TRA-retrofocusing performance in shallow water, and recent performance findings for stationary horizontal TRAs[7] parallel those in previous matched-field processing (MFP) studies.[8–10] This paper extends these stationary-array studies to include array motion. It reports the results of a theoretical and computational investigation into how array motion and orientation influence TRA retrofocusing in shallow ocean environments. The main finding is that the performance of a towed horizontal TRA degrades more rapidly with towing speed than with an equivalent (but clearly idealized) towed vertical array.

Towed arrays generally consist of a line, or multiple lines, of hydrophones housed in a neutrally buoyant, acoustically transparent hose(s). Such arrays are commonly used in military sonar systems and for marine seismic exploration. The shape of a towed array is dynamic, and is influenced by

[a)]Portions of this work were presented at the 143th ASA meeting, Fort-Lauderdale, FL, December 2001.
[b)]Author to whom correspondence should be addressed. Electronic mail: drd@engin.umich.edu

changes in the tow ship's course and speed, and by shear and currents in the water column. For many signal processing techniques (for instance, MFP), it is important to determine the relative positions of the sensors (or, equivalently, the shape of the array) in order to maximize the array-processing gain. However, the time-reversal process is robust with respect to array shape variation provided the shape does not change between the signal reception and time-reversed transmission.[1,11] Therefore, knowledge of the exact shape of the TRA is not required, an important advantage over conventional beamforming and array-processing techniques. Although other types of array motion are possible, for simplicity and clarity, the case of a TRA having uniform motion (e.g., the array shape remains constant during the recording and broadcast) responding to a stationary point source is covered here.

Unlike source motion[12] which only influences the source-to-array propagation, array motion affects the time-reversal process twice. First, during the forward propagation when the array is passive and listens, each element of the TRA acts as a moving receiver. Second, during the backward propagation, when the array transmits and plays a time-reversed version of the recorded signal, each element of the TRA acts as a moving source. Prior work on the Doppler shifted field from a point source in an oceanic waveguide[13] showed that the source/receiver dynamics are inherently nonreciprocal, with a moving source entering the formulation differently from the way a moving receiver does. Therefore it can be anticipated that constant TRA motion will not have equal and opposite effects during the forward and backward propagation steps.

The remainder of this paper is divided into three sections. In the next section, closed-form analytical results for the retrofocus field created by a TRA moving uniformly in a waveguide are derived from a normal-mode formulation of the Doppler-shifted field from a moving point source.[13] The influence of the array configuration on the performance of the time-reversal process is discussed for both the case of an idealized moving vertical TRA and a moving horizontal endfire array. In the third section, the effect of array motion is evaluated quantitatively through numerical simulations for various source-array ranges and both array configurations. A summary of findings and conclusions drawn from this study are presented in the final section.

II. THEORETICAL ANALYSIS OF MOVING TRA PERFORMANCE IN A WAVEGUIDE

A. Normal-mode formulation of the Doppler effect in an oceanic waveguide

The formal development starts from prior spectral and modal representations of the Doppler-shifted field in ocean waveguides for the case of a moving source and a moving receiver when both move with uniform (i.e., constant velocity) horizontal speeds.[13] In the moving receiver's frame of reference, both source and receiver motion lead to Doppler effects, while only the receiver motion affects the depth-dependent Green's function.[13] In this formulation it is assumed that an acoustic source is moving horizontally at a

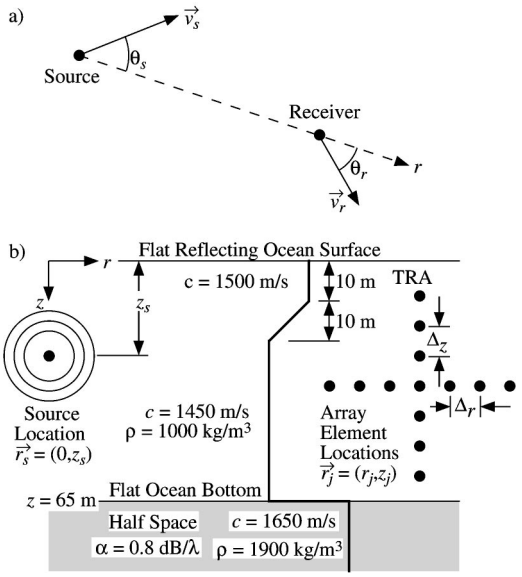


FIG. 1. (a) Top view of the moving source and receiver geometry for use with Eq. (1). (b) Side view of the computational range-independent shallow water sound channel. Environmental parameters used for the simulations are specified. The speed of sound profile is indicated by the heavier line.

constant depth z_s in a plane stratified ocean with velocity vector $\vec{v}_s$. The source is initially at the origin so that the vector position of the source is $\vec{r}_s = (\vec{v}_s t, z_s)$ in cylindrical coordinates. Similarly the receiver at depth z moves with velocity vector $\vec{v}_r$ and the vector position of the receiver is then $\vec{r}_r = (\vec{r}_o + \vec{v}_r t, z)$. The azimuthal angles between the velocity vectors and the radial vector connecting initial source and receiver locations are denoted θ_s and θ_r and are depicted in the Fig. 1(a). In the moving receiver's frame of reference, the normal-mode representation of the frequency spectrum is given by[13]

$$\psi(\vec{r}_o, z, \omega, \vec{v}_s, \vec{v}_r) = \frac{i}{4\rho(z_s)} \sum_n S(\Omega_n) \Psi_n(z) \Psi_n(z_s) H_o^{(1)} \times (k_n(\omega + k_n(\omega) v_r \cos\theta_r) \cdot r_o), \quad (1)$$

where ω is the frequency in the receiver's frame of reference, ψ is the recorded spectrum at the moving receiver, r_o is the horizontal source-array distance at $t=0$, $S(\Omega)$ is the spectrum of the source signal, $k_n(\omega)$ is the horizontal wavenumber of the nth mode at a frequency ω, $\Omega_n = \omega - k_n(\omega) \times (v_s \cos\theta_s - v_r \cos\theta_r)$, and $\Psi_n(z)$ is the nth modal eigenfunction which is a solution of

$$\rho(z) \frac{d}{dz}\left[\frac{1}{\rho(z)} \frac{d\Psi_n(z)}{dz}\right] + \left[\frac{\omega^2}{c(z)^2} - k_n^2\right] \Psi_n(z) = 0. \quad (2)$$

Note that because of the varying Doppler shifts for each mode, the source signal spectrum appearing in Eq. (1) contributes to ψ at multiple frequencies.

Changes in mode shapes are assumed negligible for small Doppler frequency shifts. In Eq. (1) the motion of the source and the receiver induces Doppler shifts in both temporal and spatial frequencies. The source spectrum is evaluated at the mode-specific Doppler-shifted frequency Ω_n, which depends on *both* the source and receiver velocities and relative directions. In addition, the Hankel function is evalu-

ated using a shifted frequency $\omega + k_n(\omega) v_r \cos\theta_r$ which depends *only* on the receiver velocity and relative direction.

Although Eq. (1) incorporates Doppler effects from both source and array motion, typical array towing speeds may make some of the predicted effects difficult to observe for broadband signals. To see this, consider the two factors that would lead to a nonzero difference $\psi(\vec{r}_0, z, \omega, \vec{v}_s, \vec{v}_r) - \psi(\vec{r}_0, z, \omega, 0, 0)$. Motion enters the formulation, Eq. (1), through the argument of the signal spectrum and through the Hankel function. To simplify this discussion, these factors are considered separately and it is assumed that $k_n(\omega) \approx \omega/c$, where c is a nominal sound speed and that the Doppler shift, $\Omega \mathrm{M}_a$ where M_a is the relative Mach number $(v_s \cos\theta_s - v_r \cos\theta_r)/c$, is small enough to use a local Taylor series expansion of $S(\Omega)$. With these simplifications, the spectral difference, $\Delta S = S(\Omega_n) - S(\omega)$, in the frequency bin centered on Ω_n will be proportional to $\Omega_n \mathrm{M}_a (dS/d\Omega)$. Thus, ΔS will be small when M_a is low, or the local spectral slope is small. Both of these conditions are likely to be met by short broadband pulses exchanged between a source and a towed TRA. However, for a narrow-band signal having a single spectral peak at Ω_c of width $\Delta\Omega$, the relative spectral amplitude change $\Delta S/S$ will only be small when $(\Omega_c/\Delta\Omega)\mathrm{M}_a$ is small. However, if the frequency resolution of the recording equipment, $\delta\Omega$, is larger than the Doppler shift, $\Omega_c \mathrm{M}_a$, then little signal energy may be shifted between frequency bins and the effects of source or array motion may be difficult or impossible to observe in the spectral changes. However, even when the spectral changes between static and moving source-receiver situations are small, a phase difference may still exist because of the receiver-speed dependence in the Hankel function. Using the same simplifications as above and ignoring Doppler amplitude changes in the Hankel function, a phase difference, which is independent of $S(\Omega)$, arises that can be expected to be proportional to $(\Omega r_0/c)(v_r \cos\theta_r)/c$, a linear function of frequency and initial source–receiver distance. Thus, in the time domain, receiver motion introduces a range-dependent time shift in the received signal. This time shift leads to performance degradation for a moving horizontal TRA array when its elements lie at different ranges from the source. However, slowly moving sources and receivers exchanging short signal pulses may be relatively unaffected by source and receiver motion.

Several simplifications of Eq. (1) will be used throughout this paper. First of all, θ_s and θ_r will be assumed zero, which yields the largest possible Doppler shift for fixed source and receiver speeds. This simplification constrains the overall source-array geometry to two dimensions (range and depth). Thus, only vertical and endfire TRAs are considered in the following sections. The performance of broadside horizontal towed TRAs can be anticipated to include the focal shift found below without significant Doppler phase modifications. Similarly, horizontal TRA configurations between endfire and broadside and can be handled by replacing v_r with $v_r \cos\theta_r$ in what follows. Second, only stationary sources are considered; source motion is a separate topic.[12] And, finally, the speed of the TRA is assumed to remain constant during reception and broadcast.

An adiabatic normal-mode approximation[14] (valid for a weakly range-dependent sound channel) can also be derived based on a direct extension of Eq. (1).[13] A more general theory of the Doppler effect created by a moving receiver having an arbitrary motion in an arbitrary environment requires the use of direct time-domain convolutions or field solutions,[15,16] which are the time-domain extension of the previous formulation given in Eq. (1). However, such numerical solutions are computationally intensive. Therefore, in the following section, for the analysis of a moving TRA, the normal-mode formulation for a receiver having a uniform horizontal motion and a stationary source in a range-independent oceanic waveguide will be used. This normal-mode formulation has the advantage of being easily computed and of having a direct connection to the waveguide physics.

B. Retrofocus field created by a moving TRA

In this subsection, Eq. (1) is used to determine a moving TRA's response to a broadband stationary point source. The time origin is set at the beginning of the emission of the source signal. The range axis starts from the source location and the depth axis from the surface. The jth TRA element has a horizontal uniform motion in range given by $r_j = r_0^j + v_r t$, where v_r is a nonzero constant, at a constant depth z_j. During the forward propagation, the spectrum of the recorded pressure field $P_a^j(\omega)$ at the jth element is obtained from Eq. (1) with $v_s = 0$, where the ranges are assumed large enough for the usual asymptotic form of the Hankel function to be valid:

$$P_a^j(\omega) = \frac{i e^{-i\pi/4}}{\sqrt{8\pi}\rho(z_s)} \sum_{n=1}^{N} S(\omega + k_n(\omega) v_r) \times \Psi_n(z_j)\Psi_n(z_s) \frac{e^{i k_n(\omega + k_n(\omega) v_r) r_0^j}}{\sqrt{k_n(\omega + k_n(\omega) v_r) r_0^j}}. \quad (3)$$

Once the signal is recorded by the array, it is then time-reversed. Complex conjugation in the frequency domain is equivalent to time reversal in the time domain.[1,11] To ensure causality—the signal must be completely received before it can be time reversed—a temporal delay T_c is applied to the complex-conjugated spectrum.[1,11] Therefore the causal time-reversed signal (complex conjugation is indicated by *) to be broadcast by the jth element is $[P_a^j(\omega)]^* e^{i\omega T_c}$. For simplicity, the array is assumed to have the same uniform motion during both recording and broadcast. Array motion during signal recording places each array element at a new range $r_b^j = r_0^j + v_r T_f$ at broadcast, where T_f corresponds to the time of the beginning of the TRA broadcast. A requirement for the causality time is that $T_c \geq 2T_f$.

During the broadcast, the TRA is active and each array element acts as a moving source with effective speed $-v_r$ since the backpropagation is directed toward the stationary source. Using the monopole formulation for the TRA, for M array elements and a causality time T_c,[11] the retrofocused pressure spectrum at the stationary location (r_f, z_f) is, using Eq. (1) for a moving source

$$P_f(r_f,z_f,\omega)=\frac{ie^{-i\pi/4}}{\sqrt{8\pi}}\sum_{j=1}^{M}\sum_{l=1}^{N}[P_a^j(\omega)+k_l(\omega)\cdot v_r]^*$$
$$\times\exp(i(\omega+k_l(\omega)\cdot v_r)\cdot T_c)\frac{\Psi_l(z_j)}{\rho(z_j)}$$
$$\times\Psi_l(z_f)\frac{\exp(ik_l(\omega)\cdot(r_0^j+v_rT_f-r_f))}{\sqrt{k_l(\omega)\cdot(r_0^j+v_rT_f-r_f)}} \tag{4}$$

The source spectrum term, $[P_a^j(\omega+k_l(\omega)\cdot v_r)]^*\exp(i(\omega+k_l(\omega)\cdot v_r)\cdot T_c)$, in the previous expression for $P_f(r_f,z_f,\omega)$ is a Doppler-frequency-shifted expression of the causal time-reversed signal recorded at each array element given by Eq. (3). The normal-mode formulation of the retrofocused field given by Eq. (4) is valid for a TRA having uniform motion in a range-independent sound channel. This expression will be used to evaluate the performance of the time-reversal process with a moving array for the numerical simulations in the next section. From Eqs. (3) and (4), in the modal sum over index n (associated with the forward propagation), the frequency ω is now replaced by the Doppler-shifted frequency $\omega+k_l(\omega)\cdot v_r$ when considering the expression for the Doppler-frequency-shifted received spectrum $[P_a^j(\omega+k_l(\omega)\cdot v_r)]$, where index l is associated with the backward propagation. For low Mach number, i.e., $v_r/c_0\ll 1$ where c_0 is a reference speed of sound in the water column, a first-order expansion of the modal phase of $[P_a^j(\omega+k_l(\omega)\cdot v_r)]$, given by Eq. (3), yields

$$k_n(\omega+k_l(\omega)\cdot v_r+k_n(\omega+k_l(\omega)\cdot v_r)\cdot v_r)$$
$$\simeq k_n(\omega)+\frac{v_r}{U_n(\omega)}(k_l(\omega)+k_n(\omega)), \tag{5}$$

where $U_n(\omega)=(dk_n/d\omega)^{-1}$ is the group velocity of the nth mode at angular frequency ω.

Using Eq. (5), the retrofocus pressure field is then given by

$$P_f(r_f,z_f,\omega)=\frac{1}{8\pi\rho(z_s)}\sum_{j=1}^{M}\sum_{l=1}^{N}\sum_{n=1}^{N}[S(\omega+(k_l(\omega)+k_n(\omega))\cdot v_r)]^*e^{i\omega T_c}\frac{\Psi_l(z_j)\Psi_n(z_j)}{\rho(z_j)}\Psi_l(z_f)\Psi_n(z_s)$$
$$\times\frac{\exp(-i(k_n(\omega)+[v_r/U_n(\omega)]\cdot(k_l(\omega)+k_n(\omega)))\cdot r_0^j)}{\sqrt{(k_n(\omega)+[v_r/U_n(\omega)]\cdot(k_l(\omega)+k_n(\omega)))\cdot r_0^j}}\frac{\exp(ik_l(\omega)\cdot(r_0^j+v_r(T_f+T_c)-r_f))}{\sqrt{k_l(\omega)\cdot(r_0^j+v_r(T_f+T_c)-r_f)}}. \tag{6}$$

Further simplification of Eq. (6) can be achieved once the configuration of the TRA is specified. In practical implementations of moving TRA concepts, the array is likely to be horizontal. However, vertical water-column-spanning TRAs are favored for research experiments and their performance is well understood. Therefore, simulations of a towed vertical TRA, although idealized, provide a performance gauge for towed horizontal TRAs. Thus, both orientations are discussed in the following subsections.

C. Moving vertical TRA

Prior investigations of vertical TRA performance[1–4,11] make this configuration ideal for illustrating how array motion affects the time-reversal process. Furthermore, the phenomena highlighted here are also active in the horizontal array case.

The point of departure for the vertical array analysis is Eq. (6). The mode orthonormality property for a vertical water-column-spanning array with a sufficient number of transducers[1,7] having a uniform vertical spacing Δ_z is

$$\sum_{j=1}^{M}\frac{\Psi_n(z_j)\Psi_l(z_j)}{\rho(z_j)}\simeq\frac{1}{\Delta_z}\int_0^H\frac{\Psi_n(z)\Psi_l(z)}{\rho(z)}dz=\frac{\delta_{ln}}{\Delta_z}. \tag{7}$$

Using this and neglecting the small Doppler amplitude adjustment in the cylindrical square-root-of-range spreading factor, Eq. (6) can be simplified for a vertical linear array located, at time $t=0$, at a range r_0^a:

$$P_f(r_f,z_f,\omega)$$
$$=\frac{1}{8\pi\rho(z_s)\Delta_z}$$
$$\times\sum_{n=1}^{N}[S(\omega+2k_n(\omega)v_r)]^*e^{i\omega T_c}\Psi_n(z_f)\Psi_n(z_s)$$
$$\times\frac{\exp(-ik_n(\omega)(v_rT_f+v_r(T_c-2r_0^a/U_n)-r_f))}{k_n(\omega)\sqrt{r_0^a}\sqrt{r_0^a+v_r(T_f+T_c)-r_f}}. \tag{8}$$

This expression shows that the retrofocus location (defined as the range corresponding to zero-phase in the asymptotic form of the Hankel function expansion) is different for each mode. Two terms contribute to this mode-dependent retrofocus shift relative to the source location (the range origin): (1) the geometric array translation during the recording, v_rT_f, independently of the mode number, and (2) the Doppler induced translation, $v_r(T_c-2r_0^a/U_n)$, which depends on each mode number through its group speed U_n. For a shallow water sound channel, higher-order modes are stripped due to bottom absorption or scattering, therefore only the remaining lower order modes contribute to the focus.[1,4] For the remaining lower-order modes the values of the group speed at a given frequency $U_n(\omega)$ are very close to each other. Hence,

as long as the array speed remains low, the retrofocus shift $(\Delta r_f)_{\text{vert}}$ induced by the lower-order modes will be almost independent of the mode numbers for a vertical array and proportional to the array speed v_r:

$$(\Delta r_f)_{\text{vert}} = v_r\left(T_f + T_c - \frac{2r_0^a}{U_{\text{av}}}\right), \tag{9}$$

where U_{av} corresponds to a reference group speed (typically close to the depth-averaged speed of sound in the water, c_{av}, where the lower-order modes propagate). In the case of the moving vertical TRA, the retrofocus will shift, and its amplitude will slightly decrease for moderate tow speeds (up to 10 m/s, see Sec. III B) because the extra phase term, in the modal sum of Eq. (8), due to the Doppler effect $(k_n v_r/U_n)r_0^a$ is nearly the same for each array element. Thus, the Doppler effect should have little overall impact on the performance of a moving vertical TRA beyond the horizontal shift induced by array translation, given by Eq. (9). This horizontal shift could be compensated using techniques based on the waveguide invariant,[17,18] if a quantitative prediction of its value is available.

Overall, Eq. (8) suggests that (*i*) for moderate towing speeds the retrofocus will merely be shifted and not changed in size, (*ii*) the sensitivity TRA performance to array motion will increase when there is wider variation in U_n, and (*iii*) the sensitivity of TRA performance to array motion decreases with decreasing pulse length but does not reach zero even if S is constant. These features of a moving TRA persist when the array is horizontal. Simulations provided in the third section of this paper support these observations and clearly illustrate the theoretical retrofocus shift.

D. Moving endfire horizontal TRA

Here the moving TRA, at depth z_a, is linear, horizontal, and endfire to the source. The initial source-to-center-of-array range is r_0^a and the spacing Δ_r between the M array elements is constant for an array length of $L=(M-1)\Delta_r$. The element spacing, array length, and array depth were chosen based on suggestions in prior MFP[10] and TRA studies.[7] Quantitative discussion of these parameter choices is provided in the next section.

Unfortunately, no simplification of Eq. (6) based on mode orthogonality [see Eq. (7)] is possible for a horizontal array having elements at constant depth. However, horizontal arrays are commonly treated by separating the double modal sum in Eq. (6) into diagonal ($l=n$) and off-diagonal ($l\neq n$) terms[7,8] because the off-diagonal terms will tend to cancel with each other while the diagonal terms will tend to reinforce each other. In the case of a moving TRA, the additional phase terms due to the Doppler effect will not change the mutual cancellation of the off-diagonal terms. Thus, the influence of the Doppler effect will be evaluated only for the diagonal terms of Eq. (6). Substituting $r_0^j = r_0^a + j\Delta_r - (M+1)\Delta_r/2$, and approximating the amplitude decay associated with the square-root cylindrical range spreading by $1/(k_n r_0^j)$ for small values of $r_f - v_r(T_c+T_f)$ in Eq. (6), the following expression for the diagonal terms of the retrofocused field emerges:

$$\begin{aligned}[P_f(r_f,z_f,\omega)]_{l=n} &= \frac{1}{8\pi\rho(z_s)\rho(z_a)}\sum_{n=1}^{N}\left([S(\omega+2k_n(\omega)v_r)]^*\right.\\ &\times e^{i\omega T_c}\Psi_n(z_f)\Psi_n(z_s))\\ &\cdot\left(\frac{\exp(-ik_n(\omega)(v_rT_f+v_r(T_c-2r_0^a/U_n)-r_f))}{k_n(\omega)r_0^a}\right)\\ &\cdot|\Psi_n(z_a)|^2\cdot F_n(v_r),\end{aligned} \tag{10}$$

where

$$F_n(v_r)=\sum_{j=1}^{M}\frac{\exp(-i2k_n(v_r/U_n)\Delta_r(j-(M+1)/2))}{1+\Delta_r(j-(M+1)/2)/r_0^a}. \tag{11}$$

In Eq. (10) the first two terms in parentheses in the modal sum are similar to those in the modal sum of the retrofocused field generated by a moving vertical array at a range equal to the initial range to the center of the array r_0^a [see Eq. (8)]. Therefore, if the retrofocus exists, it will be shifted in range for increasing source speed just like the idealized towed vertical array (as shown in the previous subsection). However, each term in the modal sum is modulated by two extra factors: (1) the magnitude of the modal eigenfunctions at the array depth $|\Psi_n(z_a)|^2$, and (2) a modal phase and amplitude correction function $F_n(v_r)$. The first factor creates an additional weighting by the vertical structure of the propagating modes, independently of the array motion, and is responsible for persistent sidelobes in depth.[7] The second factor is more critical for moving TRA performance because the phase term in Eq. (11) due to the Doppler effect $(k_n v_r/U_n)r_0^j$ varies along the array length. The resulting modal factor $F_n(v_r)$ due to this difference of Doppler shift in range makes the contributions of each array element out of phase with each other at the retrofocus location (r_f,z_f). Therefore, the retrofocus field amplitude drops and the retrofocus location shifts for nonzero array-towing speed.

When the source-array range is much larger than the spatial extent of the array (i.e., $\Delta_r(j-(M+1)/2)/r_0^a \ll 1$), the denominator of $F_n(v_r)$ can be simplified, yielding

$$F_n(v_r)\simeq\frac{\sin(M\Delta_r(k_n/U_n)v_r)}{\sin(\Delta_r(k_n/U_n)v_r)}. \tag{12}$$

To evaluate this simplified form of the modal correction factor $F_n(v_r)$, knowledge of the source-array range is not required, and the other parameters are either likely to be known (M,Δ_r,v_r) or easily estimated at the TRA location (mode shape and horizontal wave number).[19,20]

For a given range-independent sound channel (i.e., U_n and k_n are known), an important parameter under the control of an array designer is $M\Delta_r$, which is approximately the same as the array length $L=(M-1)\Delta_r$. The form of Eq. (14) suggests that the larger the horizontal aperture of the array is, the more the Doppler phase term $(k_n v_r/U_n)r_0^j$ will vary along the array length. Thus, retrofocus degradation is expected to occur more rapidly for longer TRAs as the array-towing speed is increased. Hence, selecting horizontal TRA length will be a trade-off between limiting the Doppler deg-

radation in array performance (i.e., keeping the array short) and the requirement for the horizontal array to correctly sample the propagating modes (i.e., making the array long).[7,10] Those considerations will be addressed quantitatively in the next section for a specific range-independent sound channel.

Due to the weighting of the retrofocused field by the vertical modal structure, the mode $(n=p)$ having the strongest contribution in the vicinity of the initial source location (i.e., $r_f \ll 1, z_f = z_s$) is the one which maximizes the quantity $|\Psi_n(z_a)\Psi_n(z_s)|^2 \cdot e^{2r_0^a \alpha_n}$ where $\alpha_n < 0$ is the imaginary part of the horizontal wave number, related to the modal absorption. An important factor for the amplitude of the retrofocus field is then $F_{n=p}(v_r)$. A decay of 3 dB for the intensity of the retrofocus field corresponds then roughly to a speed $(v_r)_{\max}$ that is the solution of the equation $(F_{n=p}(v_r)_{\max})^2 < \frac{1}{2}$. Quantitative evaluations of this retrofocus amplitude decay and the influence of array speed will be presented in the third section and these numerical results will be compared to predictions based on Eq. (14).

One may attempt to determine a phase correction $\gamma_j(\omega)$ for each TRA element recorded signal $P_a^j(\omega)$ [see Eq. (3)] to compensate for the relative additional Doppler phase $(k_n v_r / U_n)(r_0^j - r_0^a)$. If a value of the reference group speed $U_{\text{av}} \approx c_{\text{av}}$ [see Eq. (9) and below] is known as well as an estimate $\tilde{v}_r$ of the true value of the projected receiver velocity v_r [corresponding to $v_r \cos(\theta_r)$ see Fig. 1(a) and Sec. II A], a potential expression for $\gamma_j(\omega)$ is

$$\gamma_j(\omega) = 2k_{\text{av}} \frac{\tilde{v}_r}{c_{\text{av}}} \Delta_r \left(j - \frac{M+1}{2} \right), \tag{13}$$

where $k_{\text{av}} = \omega / c_{\text{av}}$. The new time-reversed signal to be broadcast is then $[P_a^j(\omega)]^* e^{i\gamma_j(\omega)} e^{i\omega T_c}$. The new simplified modal factor, after phase correction, $\tilde{F}_n(v_r)$ [see Eq. (14)], is now

$$\tilde{F}_n(v_r) \simeq \frac{\sin(M\Delta_r((k_n/U_n)v_r - (k_{\text{av}}/c_{\text{av}})\tilde{v}_r - k_n \tilde{v}_r v_r / c_{\text{av}}^2))}{\sin(\Delta_r((k_n/U_n)v_r - (k_{\text{av}}/c_{\text{av}})\tilde{v}_r - k_n \tilde{v}_r v_r / c_{\text{av}}^2))}. \tag{14}$$

If reliable estimates c_{av} and $\tilde{v}_r$ are available, the amplitude decay induced by $\tilde{F}_n(v_r)$ is now smaller since $|(k_n v_r / U_n - k_{\text{av}} \tilde{v}_r)/c_{\text{av}}| < k_n v_r / U_n$ and since $\tilde{v}_r v_r / c_{\text{av}}^2 \ll 1$ is a second-order term in Mach number. Thus the retrofocused amplitude of the moving horizontal TRA might be made less sensitive to Doppler effects.

E. Range-dependent environments

When a TRA moves continuously at a uniform speed, its recording and broadcast positions will be different. In a range-independent environment, the resulting geometric translation causes a retrofocus translation. However, in a range-dependent environment, this translation will induce a decorrelation of the TRA's signal at the retrofocus location since the recorded signal will be rebroadcast back toward the source along different paths. This sensitivity of the time-reversal process to the TRA displacement has been previously used in ultrasonic experiments to measure the characteristics of rough surfaces.[21,22] For this application, the TRA was stationary during the recording and then displaced before the beginning of the broadcast, so array translation, and not the Doppler effect, was the main issue.

In general, range dependency in the environment will further decrease the performance of a moving array when compared to the same array operating in an equivalent range-independent environment. In the frame of a moving array, a static range-dependent environment appears as a dynamic environment and such environments are known to degrade TRA performance.[4,23] However, the effects of environmental range dependency on moving TRA performance should be small if the range direction or horizontal variations in the environment are sufficiently gradual and if the TRA displacement is small compared to the distance necessary to observe a substantial environmental change, such as might be set by internal wave fluctuations, bottom or surface roughness, or by the angle of a sloping bottom. In this case the limitations would result essentially from the Doppler effect, and the previous analysis for range-independent environments can serve as a guide for evaluating changes caused by TRA motion.

III. NUMERICAL SIMULATIONS

In this section we present numerical simulations based on a normal-mode representation of the retrofocused pressure field generated by a moving TRA, Eq. (4). The numerical simulations were done for both a moving vertical array and a moving endfire horizontal array in a generic range-independent shallow water sound channel (Fig. 1). The primary interest was to evaluate TRA performance for different source-array configurations as a function of array-towing speed.

A. Computational approach

The range-independent shallow water sound-channel has a downward refracting sound speed profile, a depth $D = 65$ m, and a sandy bottom.[14] The range coordinate, r, runs in the horizontal direction and has its origin at the stationary source location. The z coordinate has its origin at the water surface and increases downward. The depth-dependent speed of sound c, bottom layer density, and bottom attenuation are specified in Fig. 1. The normal mode program KRAKEN[24,25] was used to compute the mode shape functions, $\Psi_n(\omega)$, and horizontal wave numbers, $k_n(\omega)$, for this sound channel. The group velocity of the nth mode at the angular frequency ω, $U_n(\omega)$, could then be evaluated using a perturbation theory:[14]

$$U_n = \left(\frac{dk_n}{d\omega} \right)^{-1} \simeq \left(\frac{\omega}{k_n} \int_0^D \frac{\Psi_n^2(z)}{\rho(z) c^2(z)} dz \right)^{-1}. \tag{15}$$

For a 500-Hz source, this sound channel supports 21 modes, the first 12 of which are illustrated in Fig. 2.

TRA performance was evaluated for a source emitting a Gaussian-windowed sine pulse $S(t) = \sin(2\pi f_c t) e^{-t^2/\tau^2}$ having a center frequency $f_c = 500$ Hz and a time constant $\tau = 10/(2\pi f_c)$. The resulting pulse bandwidth (pulse duration) is $B = 258$ Hz ($T_s = 8.3$ ms), based on 99% of the signal energy spectrum (the broadcast energy).[6,26] At the receiver

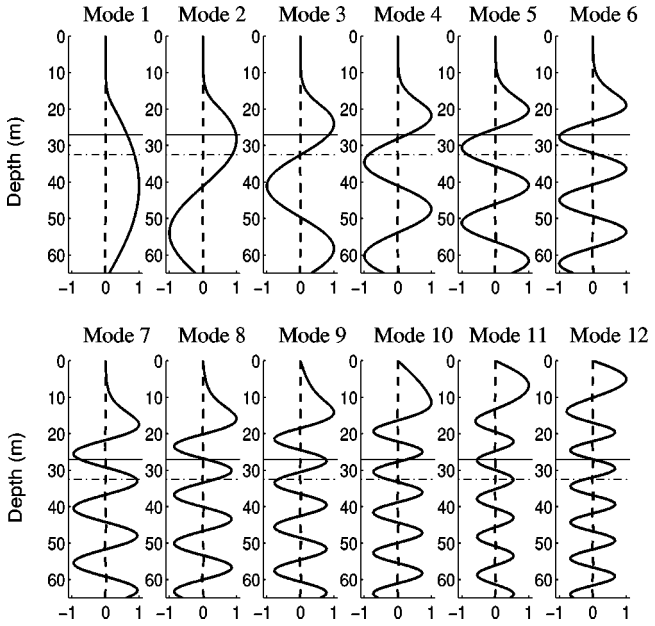


FIG. 2. First 12 modes propagating in the sound channel (see Fig. 1) at 500 Hz. Each mode is normalized by the maximum absolute value of the complex eigenfunction. For each mode the real part corresponds to the solid line and the imaginary part to the dashed line. Source depth, $z_s = 27$ m (solid line), and horizontal endfire array depth, $z_a = D/2 = 32.5$ m (dash-dot line), are also indicated.

location, time-domain signals were generated by a Fourier superposition of single-frequency results based on the output of the normal-mode code KRAKEN. The number of frequencies used for the computations was chosen so that the bandwidth B of the signal was fully covered. The period of the time window for Fourier synthesis of this Gaussian-windowed sine pulse was 1.024 s (leading to a frequency spacing between Fourier components of 0.98 Hz). These values were chosen so that the sound-channel-spread signal at the TRA location was still clearly contained in a single time window for the average source-array range of interest ($R = 2500$ m); this period was longer than the time-of-flight of the pulse along the array.

For simplicity, the array has the same uniform motion both during recording and broadcast, and its broadcast starts immediately at the end of the recording interval. Thus, $t = T_f$ specifies the time origin for the beginning of the source broadcast for all array elements. This timing assumption represents a best-case scenario when there is negligible signal processing delay at the array. Under these circumstances, there is minimal array translation between reception and broadcast and the causality delay is $T_c = 2T_f$.

In practice, determining the end of the recording window requires knowledge of the multipath spread of a signal pulse. For a shallow water sound channel, it is important that the TRA records late signal arrivals at each receiver location since they correspond to the information carried by the high-angle modes and are necessary to achieve a small retrofocus size. Here, for simplicity, the recording interval ends when 99% of the signal energy has been recorded by each array element. Thus, T_f is set by the array element furthest from the source.

In the next two subsections, the broadband performance

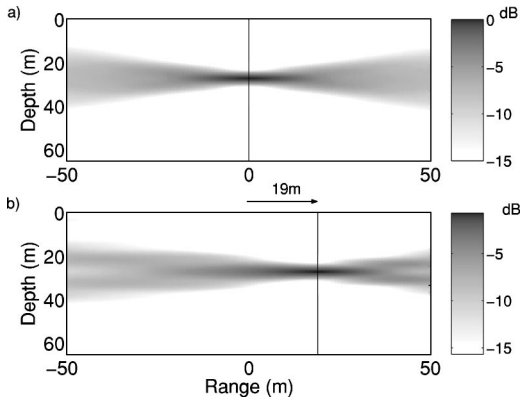


FIG. 3. Retrofocused energy distribution (in dB) in the depth versus range coordinates for an initial source/array range $r_0^a = 2500$ m. (a) Stationary TRA. (b) TRA moving away from the source at $v_r = 10$ m/s. The values were normalized by the value of the retrofocused energy by a stationary array at the initial source location ($r_s = 0$ m, $z_s = 27$ m) and the cylindrical range spreading of the energy was removed.

of a moving vertical TRA and horizontal endfire TRA are evaluated in range-independent environments.

B. Moving vertical TRA

Although towed vertical arrays are impractical, they are easily simulated and provide an intellectual bridge from stationary vertical TRAs, about which much is known, to simulations of towed horizontal TRAs which are likely to be of interest in actual underwater applications. Here, the vertical array is centered in the sound channel, has $M = 21$ elements equally spaced by $\Delta_z = 3$ m (corresponding roughly to the wavelength of the center frequency of the pulse $f_c = 500$ Hz), and spans the water column. The source emits a short Gaussian-windowed sine pulse ($f_c = 500$ Hz, $B = 258$ Hz).

Figure 3 shows the retrofocused energy distribution (in dB) in the depth versus range coordinates in the vicinity of the source for both a stationary TRA ($v_r = 0$ m/s) and a TRA moving away from the source at a speed $v_r = 10$ m/s. The array is off to the right of the figure and was initially at $r_0^a = 2500$ m at $t = 0$. The amplitude of the retrofocused energy was normalized by the value of the retrofocused energy from a stationary array at the initial source location ($r = 0$ m, $z_s = 27$ m). In both cases, the retrofocus quality is good. The main effect of the array motion is to shift the range of the retrofocus location r_f by 19 m. The moving array's retrofocus amplitude is slightly lower (by 1 dB) as well.

The evolution of the retrofocused energy (in dB), at the initial source depth $z_s = 27$ m, versus range for increasing array speed is shown in Fig. 4. The amplitude decay slightly increases with the array-towing speed due to the difference of the group speed values between the low- and high-order propagating modes. In this sound channel at 500 Hz, the first nine modes (see Fig. 2) are trapped below the mixed layer. The group speeds of these modes are slightly lower than the uniform speed of sound in the water column below the mixed layer (see Fig. 1), 1450 m/s. The higher-order modes (mode 10 and higher, see Fig. 2) have the maxima above the mixed layer and therefore have a different group speed. The

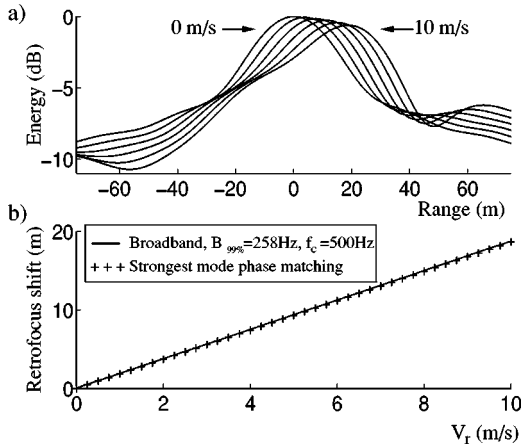


FIG. 4. (a) Retrofocused energy (in dB), at $z_f = z_s = 27$ m, versus range for increasing array speed v_r values (0, 2, 4, 6, 8, and 10 m/s from left to right). The values were normalized by the value of the retrofocused energy by a stationary array at the initial source location ($r_s = 0$ m, $z_s = 27$ m) and the cylindrical range spreading of the energy was removed. The initial source/array range is $r_0^a = 2500$ m. (b) Shift of the retrofocus range versus array speed. The solid line is the numerical results. The dashed-line is the values based on the predictions of Eq. (9).

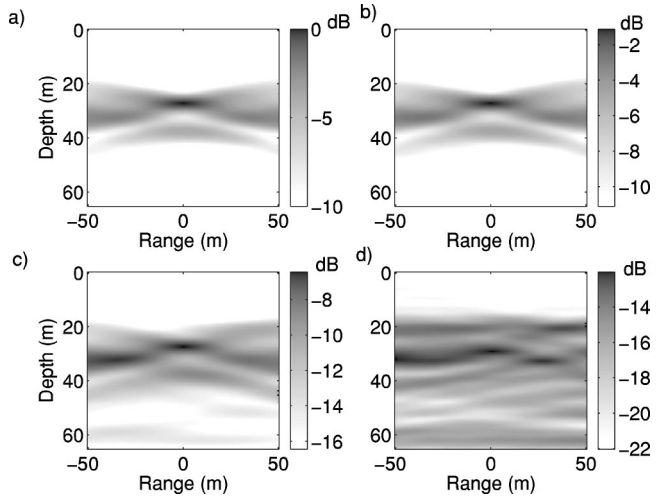


FIG. 5. Retrofocused energy distribution (in dB) in the depth versus range coordinates for an initial source/array range $r_0^a = 2500$ m. (a) Stationary TRA, (b) TRA moving away from the source at $v_r = 1$ m/s, (c) $v_r = 2.5$ m/s, and (d) $v_r = 5$ m/s. The values were normalized by the value of the retrofocused energy by a stationary array at the initial source location ($r_s = 0$ m, $z_s = 27$ m) and the cylindrical range spreading of the energy was removed.

extra phase term, in the modal sum of (8), due to the Doppler effect $v_r(k_n r_0^a / U_n)$ is proportional to the array speed v_r and to the inverse of modal group speed U_n, which has different values for the low- and high-order modes. Consequently, when v_r increases, the high-order modes are no longer in phase with the low-order modes. Thus they do not constructively contribute to the retrofocus and thereby lead to lower retrofocus amplitudes and stronger sidelobes. It should be noted that even just a half-dozen or so low-order modes still create a fine retrofocus.[1] For increasing values of v_r, the shift in range for the retrofocus is plotted in Fig. 4(b) as well as the predictions derived for this environment based on Eq. (9) using the speed of sound in the deeper part of the sound channel as an estimate of the average group speed of the low-order modes at 500 Hz, $U_{\text{av}} \simeq c_{\text{av}} = 1450$ m/s. These predicted values fit the numerical broadband results closely, and confirm the linear dependency of the retrofocus shift in range with the array speed up to $v_r = 10$ m/s.

C. Moving horizontal endfire TRA

The results presented in this subsection for a horizontal TRA are intended to be directly comparable with those in the previous subsection. Here, the moving TRA has $M = 21$ elements, is linear and horizontal, and is oriented endfire to the stationary source. The TRA is moving at constant depth $z_a = D/2 = 32.5$ m. The initial source-to-center-of-array range is $r_0^a = 2500$ m. For stationary horizontal TRAs, array element spacing (Δ_r) and array length (L) recommendations are available[7,10] based on the horizontal wave numbers of the propagating modes ($k_1, \ldots, k_N$) at a single frequency: $\pi/k_N < \Delta_r < 2\pi/(k_1 - k_N)$ and $L \geqslant 1/\min(k_{n-1} - k_n)$. The unconventional element spacing limits apply when the horizontal wave numbers of the propagating modes only occupy a finite band in spatial frequency. The array length criterion given here allows for TRAs to be a factor of 2π shorter than the usual criterion because horizontal arrays have been found to perform well in time reversal[7] and in matched field processing applications[10] when the array is just long enough to distinguish between propagating modes but not so long that the modes are uncorrelated across the array's aperture [i.e., $L \geqslant 2\pi/\min(k_{n-1} - k_n)$]. In the broadband case, the above criteria might be evaluated at the highest important frequencies in the signal spectrum. Alternatively, the theoretical results developed here [see Eq. (10)] suggest that moving horizontal TRA performance may decrease with increasing array length because larger horizontal apertures lead to more variation in the Doppler phase term $(k_n v_r / U_n) r_0^j$ across the array. Therefore, the choice of horizontal TRA parameters is a trade-off between minimizing the Doppler effect (i.e., using a short array) and correctly sampling the propagating modes (i.e., using a long array). Thus, the parameters used for the simulations presented here were $L = 500$ m $\simeq L_{\min}(700\text{ Hz})$ and $\Delta_r = 25$ m.

Figure 5 shows the retrofocused energy distribution (in dB) in the depth versus range coordinates for a stationary horizontal TRA ($v_r = 0$ m/s) and a horizontal TRA moving away from the source at three increasing speeds: $v_r = 1$, 2.5, and 5 m/s. The array is off to the right of the figure. The amplitude of the retrofocused energy was normalized by the value of the retrofocused energy at the initial source location ($r = 0$ m, $z_s = 27$ m) from a stationary array. Each frame of Fig. 5 has its own grayscale normalization.

When compared to the equivalent vertical array results, Fig. 3, the horizontal TRA retrofocus degrades more quickly with increasing array speed. As before, the retrofocus location shifts as well. For array-towing speed higher than $v_r = 2.5$ m/s the amplitude at the shifted retrofocus location is down 6 dB [Fig. 5(c)]. For the highest speed [$v_r = 5$ m, Fig. 5(d)], the retrofocus energy is spread in the water column because the Doppler effect ruins the phase of the M back-propagated signals from the array elements, thereby prevent-

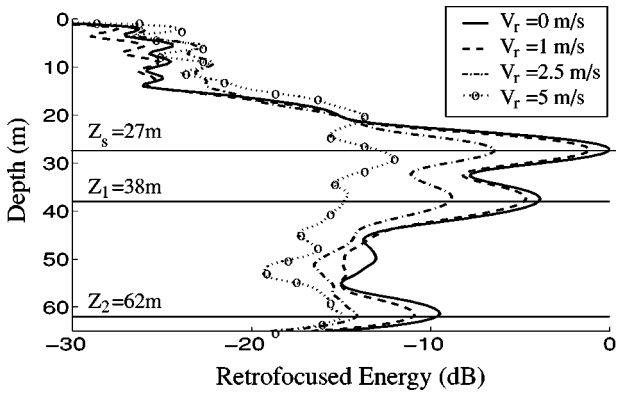


FIG. 6. Retrofocused energy distribution (in dB) versus depth, at the effective retrofocus shifted location r_f in range. Same parameters and normalization as in Fig. 5. The source depth $z_s=27$ m is indicated as well as the estimated theoretical depth $z_1=38$ m and $z_2=62$ m of the sidelobes.

ing constructive interference at a well-defined retrofocus spot at the initial source depth.

This motion-induced loss of performance is further illustrated in Fig. 6, which shows retrofocused signal energy profiles through the depth at the shifted retrofocus locations for the towed endfire horizontal TRA. These numerical results show two prominent sidelobes at $z_1=37.5$ m and $z_2=61.5$ m approximately. These sidelobes exist independently of the array speed. For a stationary horizontal TRA, the modal eigenfunctions at the array depth create an additional weighting [see Eq. (10)] by the vertical structure of the propagating modes. In a Pekeris waveguide, this weighting is responsible for two persistent TRA-produced sidelobes through the depth.[7] The depths of these two sidelobes, z_1 and z_2, were determined to be $z_1=2D-2z_a-z_s$ and $z_2=2z_a-z_s$ where, if $-D\leqslant z_i<0$, then z_i is replaced by $|z_i|$, or if $D<z_i\leqslant 2D$, then z_i is replaced by $2D-z_i$. In this sound channel, the low-order modes are trapped below approximately $z\approx 15$ m due to the downward refracting profile (see Fig. 2). Thus, the effective oceanic waveguide for low-mode acoustic propagation is between $z=15$ m and the bottom boundary at $D=65$ m; that is, the apparent water depth[3] is $D_a=D-15$ m$=50$ m. Using D_a for D to estimate the sidelobe depths yields $z_1=62$ m and $z_2=38$ m, which agree nicely with the simulation results. Thus, sidelobe formation at these depths is merely the result of horizontal TRA depth and water column structure; it is not caused by array motion.

D. Influence of the moving TRA configuration on the retrofocused energy

This subsection provides a direct comparison of the TRA retrofocused energy for a variety of array configurations. Since a horizontal focal shift is produced by all towed TRAs, the results are presented in terms of the peak signal energy found at the shifted retrofocus location. The main findings are plotted in Fig. 7, which shows the maximum retrofocused energy value (in dB) on a horizontal line through the source depth versus array-towing speed. In each case, the TRA has $M=21$ elements and the energy values are normalized by the value of the retrofocused energy produced at the initial source location ($r=0$ m, $z_s=27$ m) by an equivalent stationary array. The eight cases shown include results from vertical, horizontal, and tilted arrays.

The idealized water-column-spanning vertical TRA performs best [see Fig. 7(a) where $\Delta_z=3$ m and $r_a^0=2500$ m]

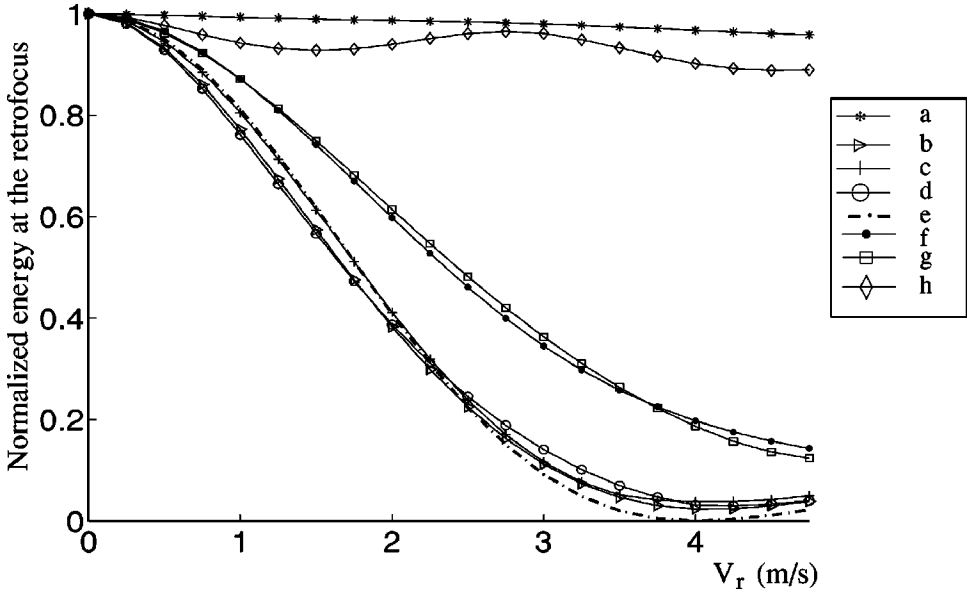


FIG. 7. Maximum retrofocused energy distribution in range (in dB) at $z_f=z_s$ versus array speed v_r for several array configurations. The initial source/array range is r_a^0. The TRA has 21 elements in each case. (a) Vertical array, $\Delta_z=3$ m, $r_a^0=2500$ m, the TRA spans the whole water. (b) Horizontal endfire array, $\Delta_r=25$ m, $z_a=D/2=32.5$ m, $r_a^0=2500$ m. (c) Same as (b) but $r_a^0=1500$ m. (d) Same as (b) but $r_a^0=5000$ m. (e) Theoretical predictions based on Eq. (14) for $\Delta_r=25$ m and $M=21$, for the strongest excited mode $n=2$. (f) Tilted array, $\Delta_z=3$ m, $\Delta_r=25$ m, $r_a^0=2500$ m, the TRA spans the whole water column in depth and has the same horizontal length than the horizontal endfire array $L=500$ m. (g) Tilted array, $\Delta_z=2.08$ m, $\Delta_r=25$ m, $r_a^0=2500$ m, the TRA spans $\frac{2}{3}$ from of the water column span (down to 43.7 m) and has the same horizontal length as the horizontal endfire array $L=500$ m. (h) Tilted array, $\Delta_z=2.6$ m, $\Delta_r=2$ m, $r_a^0=2500$ m, the TRA spans 80% of the water column (from 2 to 54 m) and has a short horizontal length $L=40$ m. For each TRA configuration, the values were normalized by the value of the retrofocused energy given by the stationary array case at the initial source location ($R=0$ m, $z_s=27$ m), and the cylindrical range spreading of the energy was removed.

because all the source-element ranges are the same leading to uniform Doppler-generated phase across the array. Therefore, the time-reversed signal generated by the M elements interferes constructively in the shifted retrofocus zone and the retrofocused energy decays mildly with increasing array speed v_r, due to increasing loss of high-order-mode contributions. At the opposite end of the performance ranking, a moving horizontal endfire TRA ($\Delta_r = 25$ m, $z_a = D/2 = 32.5$ m) performs worst. The energy retrofocused by such a moving horizontal endfire TRA decays rapidly with increasing speed v_r because the Doppler phase for each element is different. Moreover, this array's performance is similar for different initial source-array ranges [see Figs. 7(b), (c), and (d) for $r_a^0 = 2500$, 1500, and 5000 m, respectively] and this simulated performance is well approximated by the simple form $F_p(v_r)^2$ from Eq. (14) for a mode index $p = 2$ corresponding to the strongest excited mode [Fig. 7(e)]. Recall that mode p is defined as the mode maximizing the quantity $|\Psi_n(z_a)\Psi_n(z_s)|^2 \cdot e^{2r_0^a \alpha_n}$ at the center frequency of the pulse $f_c = 500$ Hz, where $\alpha_n < 0$ is the modal absorption (see Sec. II D). For the chosen computational environment, the strongest excited mode for a source depth $z_s = 27$ m and an array depth $z_a = 32.5$ m is mode 2 (see Fig. 2).

Although the idealized vertical configuration yields the best performance when the TRA is moving, it is not a practical possibility for a towed array. A tilted array, however, is more plausible and results for three tilted array configurations are shown in Fig. 7 as curves (f), (g), and (h). For each tilted array configuration, the geometric characteristics (Δ_z, Δ_r and $r_a^0 = 2500$ m) were chosen so that the propagating modes have a unique representation.[10] Figure 7(f) is produced by a tilted TRA that spans the water column with a horizontal aperture of 500 m (equal to the previous endfire horizontal array aperture). Figure 7(h) is produced by a TRA of similar horizontal aperture that only spans $\frac{2}{3}$ of the water column (from near the surface to 43.7-m depth). The performance of these two tilted TRAs with increasing tow speed is similar and intermediate between that of the vertical and the horizontal TRAs. Figure 7(h) shows the performance of a more-nearly-vertical tilted TRA that spans 80% of the water column (from near the surface to 54-m depth which keeps the array away from the bottom) and has a short horizontal length $L = 40$ m ($\Delta_z = 2.6$ m, $\Delta_r = 2$ m, $r_a^0 = 2500$ m). The performance for this moving TRA is less sensitive overall to array motion and the amplitude decay is closer to the limit set by the vertical array configuration. However, the maximum retrofocused signal energy does not decay uniformly with increasing v_r, but rather has an oscillatory behavior. Analytical predictions for a moving tilted array are difficult to establish because of the coupling between range and depth coordinates in Eq. (6).

The results summarized in Fig. 7 impact two important array design characteristics. First, a moving TRA should have a short horizontal aperture so that its performance is not severely limited by the Doppler effect. Second, a moving tilted TRA need not sample the whole water column depth to produce an acceptable retrofocus. These are both advantages for practical situations because shorter arrays are easier to deploy and retrieve, and the lower extremity of a moving tilted array can be kept away from obstacles on the ocean floor for only a minor performance penalty.

IV. SUMMARY AND CONCLUSIONS

Practical underwater applications of acoustic time reversal may require moving transducer arrays. Array motion influences the reception and broadcast steps of acoustic time reversal in ways that are not equal and opposite. Thus, motion of a time reversing array (TRA) leads to performance losses that increase with array-towing speed and depend strongly on array geometry.

The research reported here leads to four main conclusions. First, all towed TRAs will produce a retrofocus shift related to their translation between signal reception and broadcast times. In range-dependent sound channels this focal shift is likely to be accompanied by a degradation in retrofocus amplitude compared to an equivalent stationary-array and range-independent-sound channel situation. This focal shift can be eliminated if the array broadcasts from the same position(s) as it receives (this might be possible if the TRA is towed in a circle). Second, an idealized vertical water-column spanning TRA is less sensitive to towing speed in the shallow ocean than an equivalent horizontal TRA, because the vertical TRA's elements see a uniform Doppler phase. Third, the performance of a towed endfire horizontal TRA suffers compared to a vertical array because of the additional mode-shape weighting in its output, and because each of its elements has a different Doppler phase when it is towed. However, shorter horizontal arrays suffer less than longer arrays because the Doppler phase differences increase with array length. Furthermore, as the array is rotated away from endfire toward broadside, its range-direction aperture decreases and the projected velocity component decreases as well, thus endfire is likely to produce the worst-case performance. And, finally, a tilted TRA may prove to be the most versatile in any particular towed TRA application because the addition of vertical aperture helps overcome the mode-weighting problems of purely horizontal arrays and because even a little vertical aperture enables a shorter array to distinguish the propagating modes.

ACKNOWLEDGMENTS

The research was sponsored by the Ocean Acoustics Program of the Office of Naval Research under Grant No. N00014-96-1-0040.

[1] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. R. Jackson, "Phase-conjugation in the ocean: Experimental demonstration of an acoustic time reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).
[2] P. Roux and M. Fink, "Time reversal in a waveguide: Study of the temporal and spatial focusing," J. Acoust. Soc. Am. **107**, 2418–2429 (2000).
[3] S. Kim, G. F. Edelmann, W. A. Kuperman, W. S. Hodgkiss, and H. C. Song, "Spatial resolution of time-reversal arrays in shallow water," J. Acoust. Soc. Am. **110**, 820–829 (2001).
[4] M. R. Dungan and D. R. Dowling, "Computed narrowband time-reversing array retrofocusing in a dynamic shallow ocean," J. Acoust. Soc. Am. **107**, 3101–3112 (2000).
[5] S. Khosla and D. Dowling, "Time-reversing array retrofocusing in noisy environments," J. Acoust. Soc. Am. **109**, 538–546 (2001).
[6] K. Sabra, S. Khosla, and D. Dowling, "Broadband time-reversing array

retrofocusing in noisy environments,'' J. Acoust. Soc. Am. **111**, 823–830 (2002).
[7] M. R. Dungan and D. R. Dowling, ''Orientation effects on linear time-reversing array retrofocusing in shallow water,'' J. Acoust. Soc. Am. **112**, 1842–1852 (2002).
[8] C. W. Bogart and T. C. Yang, ''Source localization with horizontal arrays in shallow water: Spatial sampling and effective aperture,'' J. Acoust. Soc. Am. **96**, 1677–1686 (1994).
[9] T. C. Yang and T. Yates, ''Matched-beam processing: Application to a horizontal line array in shallow water,'' J. Acoust. Soc. Am. **104**, 1316–1330 (1998).
[10] S. L. Tantum and L. W. Nolte, ''On array design for matched-field processing,'' J. Acoust. Soc. Am. **107**, 2101–2111 (2000).
[11] D. R. Jackson and D. R. Dowling, ''Phase conjugation in underwater acoustics,'' J. Acoust. Soc. Am. **89**, 171–181 (1991).
[12] K. G. Sabra and D. R. Dowling, ''Broadband performance of a time-reversal array with a moving source,'' J. Acoust. Soc. Am. (to be published).
[13] H. Schmidt and W. A. Kuperman, ''Spectral and modal representations of the Doppler-shifted field in ocean waveguides,'' J. Acoust. Soc. Am. **96**, 386–395 (1994).
[14] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustis* (AIP, Woodbury, NY, 1994), pp. 41, 272–275, 314–315, 320–323.
[15] P. M. Morse and H. Fesbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 837, Eq. (7.3.5).
[16] K. V. Rao, T. M. Michaud, and H. Schmidt, ''Doppler shifts in underwater acoustic using field solutions,'' in OCEANS-91, IEEE Proc., Honululu, HI, 1991.
[17] H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, ''A time-reversal mirror with variable range focusing,'' J. Acoust. Soc. Am. **103**, 3234–3240 (1998).
[18] W. S. Hodgkiss, H. C. Song, W. A. Kuperman, and T. Akal, ''Long-range and variable focus phase-conjugation experiment in shallow water,'' J. Acoust. Soc. Am. **105**, 1597–1604 (1999).
[19] P. Hursky, W. S. Hodgkiss, and W. A. Kuperman, ''Matched-field source localization using data-derived modes,'' J. Acoust. Soc. Am. **107**, 2867 (2000).
[20] T. B. Neilsen and E. K. Westwood, ''Extraction of acoustic normal mode depth functions using vertical line array data,'' J. Acoust. Soc. Am. **111**, 748–756 (2002).
[21] J. H. Rose, M. Bilgen, P. Roux, and M. Fink, ''Time-reversal mirrors and rough surfaces: Theory,'' J. Acoust. Soc. Am. **106**, 716–723 (1999).
[22] P. Roux, M. Fink, and J. H. Rose, ''Time-reversal mirrors and rough surfaces: Experiment,'' J. Acoust. Soc. Am. **106**, 724–732 (1999).
[23] J. S. Kim, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, G. F. Edelmann, and T. Akal, ''A method for robust time-reversal focusing in a fluctuating ocean,'' in OCEANS. MTS-IEEE Conference and Exhibition (2001), Vol. 2, pp. 793–798.
[24] M. Porter and E. L. Reiss, ''A numerical method for ocean-acoustic normal modes,'' J. Acoust. Soc. Am. **76**, 244–252 (1984).
[25] M. Porter, *The* KRAKEN *normal mode program*, Rep. SM-245, SACLANT Undersea Research Centre, La Spezia, Italy (1991).
[26] J. G. Proakis, *Digital Communications* (McGraw-Hill, New York, 1995).

Evaluation of the integrals of target/seabed scattering using the method of complex images

John A. Fawcett
DRDC—Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

Raymond Lim
Coastal Systems Station, Code R21, 6703 West Highway 98, Panama City, Florida 32407

(Received 9 August 2002; revised 12 May 2003; accepted 23 June 2003)

In this paper the integrals which arise in target/seabed scattering problems are examined. These are the integrals which express the propagation of a spherical harmonic term in a fluid above or below a half-space and the integrals which compute the conversion coefficients of outgoing spherical harmonics into incident spherical harmonics after a seabed reflection. A very efficient and accurate method of computing these integrals, using the method of complex images, is derived. Numerical comparisons with the exact integral expressions show the accuracy of this approach. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1600726]

PACS numbers: 43.30.Gv, 43.20.El [RAS]

I. INTRODUCTION

In the T-matrix[1–5] method for solving scattering problems, the incident and scattered wave fields are expanded in terms of a basis set with regular and outgoing functions. The set of spherical harmonics is often used. When an object is located on or above an interface, a portion of the scattered field from the target will reflect off the interface before propagating to a receiver. Also, the wave field reflected from the interface will impinge upon the object and rescatter. The wave field which results from the reflection of a spherical harmonic off an interface can be expressed in terms of a wave number integral. The spherical harmonic expansion coefficients for this reflected wave field, in terms of the incoming (regular) spherical harmonic functions, can also be expressed in terms of wave number integrals.[2–5]

For a buried target, a portion of the scattered field will reflect off the seabed interface above the target and be incident upon the target. The integrals which express the spherical harmonic coefficients for this wave field are analogous to those for the case of the target in the water column. The scattered field which is transmitted back into the water column from the seabed can also be expressed as a wave number integral. As will be shown, this type of integral is significantly different in character to the other integrals.

The various wave number integrals will be discussed in detail in the following section; it will be clear that care must be taken in their numerical evaluation; the integrands may be oscillatory and the range of integration may need to be large. The method of complex images has been used recently in atmospheric[6] and underwater acoustics[7,8] and is frequently used in the electromagnetic literature (for example Refs. 9–11) as a means of accurately approximating the wave number integral expression for the reflected wave field from an interface using a few free-space point sources with complex amplitudes and positions. In this paper, we will first use the complex-image method to approximate the integrands of the wave number integrals. The integrals which result with this complex image approximation can then be evaluated analytically and the reflection or transmission of a spherical harmonic from or through an interface characterized by simple expressions. In the case of the wave number integral for the transmission of the scattered field from a buried target into the water column, the method of complex images needs to be modified to provide an efficient representation.

A large percentage of the computational time in setting up the scattering equations for an object/interface scenario and in propagating the computed scattered field to the receiver is spent in the computation of these various wave number integrals. Thus, the closed-form expressions for these various integrals which are obtained in this paper, through the use of the method of complex images, provide an approach for significantly improving computational efficiency. A number of numerical examples will be presented illustrating the accuracy of the method.

II. THEORY

A. The wave number integrals for scattering by an object near a seabed

Previous authors[2–5] have described modified T-matrix methods for modeing scattering from objects near a seabed interface (either proud or buried). Much of the theory in this section has been described in the references, but is outlined here in order to define our notation and describe the integrals that are of interest to us. In concept, the modification to the standard free-space T-matrix method is fairly straightforward. Consider the case where the incident field is characterized by ''incoming'' spherical harmonic functions

$$p^{\mathrm{inc}}=\sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty} p_{mn}^{\mathrm{inc}} P_n^m(\cos\phi) j_n(kR) e^{im\theta}, \tag{1}$$

where (j_n, h_n) denote spherical harmonic Bessel functions, P_n^m are the Legendre functions, k is the wave number, and (R, ϕ, θ) are the radial and angular coordinates of the standard spherical coordinate system, with ϕ being measured off

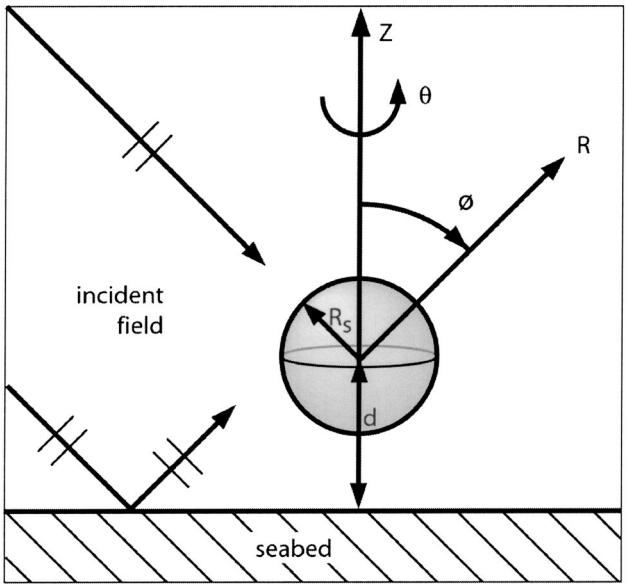


FIG. 1. A schematic drawing showing an example geometry of a sphere above the seabed.

the vertical z axis (Fig. 1). In this paper the usual Legendre functions have been scaled by

$$\sqrt{(2n+1)\frac{(n-m)!}{(n+m)!}}(-1)^n. \tag{2}$$

Then, in the standard T-matrix approach we can write

$$p_{mn}^{sc} = T p_{mn}^{\text{inc}}, \tag{3}$$

where p_{mn}^{sc} are the coefficients of the outgoing basis functions $h_n(kR)P_n^m(\cos\phi)$. In the case that the target is azimuthally symmetric, then we can apply a T-matrix for each azimuthal order m. Furthermore, for a sphere the T-matrix is simply diagonal. A particularly simple case is that of a rigid sphere, for which we can explicitly write the diagonal element for the (m,n) harmonic as

$$T_{nk}^m = -\frac{j_n'(kR_S)}{h_n'(kR_S)}\delta_{nk}, \tag{4}$$

where R_S is the radius of the sphere. The case of an elastic-shelled sphere is conceptually the same, but the expression for the diagonal elements is more complicated. For the simple case of a sphere, the T-matrix method is equivalent to simply using the separation of variables in spherical coordinates and applying the appropriate boundary conditions to the spherical harmonics at the sphere's shell or solid boundaries (see, for example, Ref. 12). In the numerical examples, for simplicity, we will consider a rigid sphere. However, it should be emphasized, the discussion also applies equally to more general target shapes where the T-matrix is full.

Let us now consider the case of the object just above a seabed. In the case of a sphere near a pressure release or rigid surface, one could simply start with a coupled system of spheres (the target and its reflected image) and solve the scattering problem as is done in, for example, Refs. 13–15. The more general case of a seabed interface whose reflection coefficient varies with wave number (the cases of a pressure release or rigid interface are subcases of this with the reflection coefficient being -1 and 1, respectively) is more problematic because wave number integrals arise that cannot be equated with real images of the target. In this paper, we show how to deal with these integrals efficiently using the method of complex images.

In order to derive the wave number integrals of interest, we first consider a single outgoing spherical harmonic of the field scattered from a target, $h_n(kR)P_n^m(\cos\phi)$. This function has the plane-wave decomposition, for $z<0$

$$h_n(kR)P_n^m(\cos\phi) = i^{n-m}\int_0^\infty \frac{hJ_m(hr)e^{-i\gamma z}}{k\gamma}P_n^m\left(\frac{\gamma}{k}\right)dh, \tag{5}$$

and for $z>0$

$$h_n(kR)P_n^m(\cos\phi) = i^{n-m}\int_0^\infty \frac{hJ_m(hr)e^{i\gamma z}}{k\gamma}P_n^m\left(-\frac{\gamma}{k}\right)dh, \tag{6}$$

where $\gamma \equiv \sqrt{k^2-h^2}$ and J_m is the cylindrical Bessel function of order m. Thus, the upgoing wave field resulting from the reflection of the (v,m) spherical harmonic off the seabed is given by

$$U_{vm}(r,z) = i^{v-m}\int_0^\infty J_m(hr)e^{2i\gamma d}R(h)P_v^m\left(\frac{-\gamma}{k}\right)\frac{e^{i\gamma z}}{k\gamma}hdh, \tag{7}$$

where d is the vertical offset of the center of the object from the seabed and $R(h)$ is the seabed reflection coefficient as a function of the horizontal wave number h. Further, using the expansion

$$J_m(hr)e^{i\gamma z} = \sum_{u=m}^{\infty} i^{u-m}P_u^m(\cos\phi)P_u^m\left(\frac{\gamma}{k}\right)j_u(kR), \tag{8}$$

we obtain an expression for the conversion coefficient of the vth outgoing harmonic into the vth incoming harmonic $[j_u(kR)P_u^m(\cos\phi)]$

$$C_{uv}^m \equiv i^{v+u-2m}\int_0^\infty P_u^m\left(\frac{\gamma}{k}\right)P_v^m\left(\frac{\gamma}{k}\right)\frac{R(h)e^{2i\gamma d}}{k\gamma}hdh \tag{9}$$

(note the symmetry of the expression).

It can be shown that in the presence of the seabed the modified T-matrix relation (for each azimuthal order) is

$$p_{mn}^{sc} = (T^{-1}-C^m)^{-1}p_{mn}^{\text{inc}}, \tag{10}$$

where the incident coefficients now also account for a seabed reflection. One can derive this equation by formally performing a sum of all the target/seabed interactions (see, for example, Ref. 5). For the case of a plane wave incident with vertical angle ϕ_{inc}, then

$$p_{mn}^{\text{inc}} = i^n[P_n^m(\cos\phi_{\text{inc}}) + R(\phi_{\text{inc}})P_n^m(-\cos\phi_{\text{inc}}) \times \exp(2i\cos\phi_{\text{inc}}d)]. \tag{11}$$

Once the p_{mn}^{sc} have been determined, the scattered field in the waveguide is given by

$$p^{sc}(x,y,z)=\sum_{m=0}^{\infty}\epsilon_m\sum_{n=m}^{\infty}p_{nm}^{sc}[P_n^m(\cos\phi)h_n(kR)+U_{nm}(r,z)]\cos(m(\theta-\theta_s)), \tag{12}$$

where $U_{nm}(r,z)$ is given by Eq. (7) and $\epsilon_m=2$ for $m>0$ and $\epsilon_0=1$. In our computations $\theta_s=0$. We have assumed azimuthal symmetry in the problem in order to use the cosine expansion in terms of θ.

For the case of a buried target, the expressions for the conversion integrals are the same as Eq. (9) except now the reflection coefficient is in terms of the seabed to water reflection and the vertical wave number γ is that for the seabed. The integral expression for the transmission of a spherical harmonic through the seabed/water interface is given by

$$\overline{U}_{nm}(r,z)=i^{n-m}\int_0^{\infty}J_m(hr)e^{i\gamma_2 d}\Gamma(h)\times P_n^m\left(\frac{\gamma_2}{k}\right)\frac{e^{i\gamma_w(z-d)}}{k\gamma_2}h\,dh, \tag{13}$$

where $\Gamma(h)$ is the transmission coefficient and γ_w and γ_2 are the vertical wave numbers for the water and seabed, respectively. This integral is significantly different than that of Eq. (7) because of the fact that there are two different vertical wave numbers. As will be discussed later, this has implications in obtaining an efficient complex image representation.

There are two basic types of wave number integrals which need to be evaluated for the target/seabed scattering problem: (1) the integrals for computing the modified T-matrix system [Eq. (10)] and then once the scattered field has been computed, (2) the integrals for propagating the scattered field to the receiver (if the incident field was taken to be that from a point source, then the incident field would also involve integrals of this same form).

Numerically, these can be difficult integrals to evaluate. An analysis of the integrands of Eq. (9), Eq. (7), and Eq. (13) shows that a problem is the behavior of the Legendre function $P_n^m(\gamma/k)$ for large h. In particular, for $h\to\infty$, $\gamma/k\to ih$ and $P_n^m(z)$ is the order of z^n. Thus, it can be seen that for increasing n it is necessary to take the upper limit of integration larger and larger, before the exponential term will eventually cause the integrand to decay. In fact, it can be seen that the situation is particularly bad for the integrals of Eq. (9), where there are products of Legendre functions. Also, the discretization must be sufficiently fine to account for the behavior of the reflection coefficient $R(h)$. The propagation integrals of Eq. (7) and Eq. (13) can sometimes be even more problematic due to the highly oscillatory behavior of $J_m(hr)$ for large r.

Thus, although it is certainly possible to compute these integrals numerically, the integrals can be time consuming and may involve convergence tests with respect to the upper integration limit and the discretization size. In the section below, we show that using the method of complex images to replace the reflection (or transmission) coefficient in these wave number integrals by a finite approximating sum, the resulting integrals can be evaluated analytically and simple expressions obtained.

B. Evaluation of the wave number integrals using the method of complex images

In order to evaluate the wave number integrals previously described, we start by using the complex-image approximation to the reflection coefficient,[6–11] which, in the wave number domain, is an approximating set of complex exponentials such that

$$R(h)\approx\sum_{q=1}^{Q}a_q\exp(\gamma(h)\alpha_q), \tag{14}$$

where both a_q and α_q are complex. This corresponds to approximating the half-space Green's function by

$$G(x,y,z;x_s,y_s,z_s)=\frac{e^{ikD}}{4\pi D}+\sum_{q=1}^{Q}a_q\frac{e^{iKD_q}}{4\pi D_q}, \tag{15}$$

where D is the direct distance and

$$D_q\equiv\sqrt{(x-x_s)^2+(y-y_s)^2+(z+z_s-i\alpha_q)^2}. \tag{16}$$

In our case z_s will be $2d$, where d is the vertical distance from the seabed to the center of the scatterer.

In this paper, we are using the complex images as a mathematical tool for efficiently evaluating wave number integrals, but they do have a physical interpretation in terms of Gaussian beams as described in, for example, Ref. 16. It is also important to note that the physics of the wave propagation is contained in the behavior of $R(h)$. Once, $R(h)$ has been replaced by an approximating sum of terms, the remainder of the derivation is simply an application of formulas for integrals and special functions. Thus, the accuracy to which the method of complex images models the wave propagation effects is basically determined by the accuracy with which it approximates $R(h)$. For the case that $R(h)$ has singular (or near-singular) behavior, then it is difficult to obtain an accurate approximation with a finite sum of terms. This is the case for an elastic seabed[7] if one wishes to model the Scholte pole, which gives rise to the Scholte interface wave. In this case, it is necessary to analytically subtract out the pole contribution to the integrand and treat the pole contribution and the remaining "smooth" portion of the integrand separately, as described in Ref. 7. We only consider fluid seabeds in this paper, so this problem will not arise in our examples. However, even for fluid seabeds, there will be regions of horizontal wave number space where the reflection coefficient may vary rapidly. For example, the variation of the reflection coefficient in the neighborhood of the seabed critical angle gives rise to the lateral or head wave.[17] The approximation error of the complex image sum to the reflection coefficient will tend to be largest (although still small) in such regions of rapid variation, and this error will translate into the spatial/time domain as an inaccuracy in the lateral wave. If necessary, this error could be reduced by improving the complex image fit to the reflection coefficient (e.g., by using more images), or it should be possible to combine an analytical expression for the lateral wave behavior with a complex-image expression for the remainder of the wave field. However, for the examples of this paper (and for the examples of Refs. 7 and 8) no special treatment of the lateral wave (or equivalently, the critical angle region of horizontal

wave number space) is used. The sum of complex images yielded an excellent approximation in both the wave number and spatial domains. It should also be emphasized that the benchmark method used in this paper is numerically exact and hence models the lateral waves, evanescent waves, etc. exactly.

It is also possible to use a sum of complex images to approximate the reflection coefficient from a layered seabed.[8,11] The method in this case is the same as for the homogeneous seabed; however, the coefficient may be more complicated as a function of wave number, thus requiring more complex images to provide an accurate approximation. Also, as is discussed later in the paper, for the homogeneous seabed the complex image parameters scale simply as a function of frequency and need only be computed at a single frequency. This is not the case for a layered seabed.

When the seabed is homogeneous, $R(h)$ has the simple form

$$R(h)=\frac{\rho_2\gamma_1(h)-\rho_1\gamma_2(h)}{\rho_2\gamma_1(h)+\rho_1\gamma_2(h)}, \tag{17}$$

where γ_1, γ_2 are the vertical wave numbers for the water and seabed, respectively. For the $q=0$ term in Eq. (14) it is convenient to choose $a_0=(\rho_2-\rho_1)/(\rho_2+\rho_1)$ and $\alpha_0=0$. This term represents the asymptotic behavior of the reflection coefficient for $h\to\infty$. For the case of a layered seabed, the reflection coefficient is more complicated and is obtained by solving a system of equations to satisfy the continuity conditions at each layer interface.[17] The asymptotic behavior of the coefficient is once again modeled by choosing $a_0=(\rho_2-\rho_1)/(\rho_2+\rho_1)$, but now ρ_2 refers to the density of the uppermost sediment layer.

Considering just one of the terms in the expression for $R(h)$ in Eq. (7), we can now evaluate this integral term for U_{nm} [the reflected (n,m) spherical harmonic] as a simple analytic expression

$$U_{nm}(r,z;q)=i^{n-m}(-1)^{n+m}a_q\int_0^\infty J_m(hr)e^{2i\gamma d}e^{\alpha_q\gamma}\times P_n^m\left(\frac{\gamma}{k}\right)\frac{e^{i\gamma z}}{k\gamma}h\,dh \tag{18}$$

$$=(-1)^{n+m}a_q h_n(kD_q)P_n^m(\cos\phi_q), \tag{19}$$

where $\cos\phi_q\equiv(z+2d-i\alpha_q)/D_q$. This follows from using the representation, Eq. (6), and considering a complex shift to the z coordinate. Also, we have used the fact

$$P_n^m(-z)=(-1)^{n+m}P_n^m(z). \tag{20}$$

Summing Eq. (19) over the Q complex images then yields the expression for the wave field resulting from the reflection of the (n,m) outgoing spherical harmonic

$$U_{nm}(r,z)=(-1)^{n+m}\sum_{q=1}^{Q}a_q h_n(kD_q)P_n^m(\cos\phi_q). \tag{21}$$

This expression can then be used in Eq. (12) for the computation of the scattered field in the water column.

It is also known that we can write

$$h_{n_2}(kR_q)P_{n_2}^m(\cos\phi_q)=\sum_{n_1=m}^{\infty}t_{n_1}^{mn_2}j_{n_1}(kR)P_{n_1}^m(\cos\phi), \tag{22}$$

where t_u^{mv} denotes the translation coefficient relating the spherical Hankel functions in one coordinate system to the spherical Bessel functions in another coordinate system offset by a vertical translation (in this case, complex). This relation is valid for $|2d-i\alpha_q|>R$.

Thus, we can identify for the conversion coefficient of Eq. (9) that

$$C_{uv}^m=(-1)^{m+v}\sum_{q=1}^{Q}a_q t_u^{mv}(2d-i\alpha_q), \tag{23}$$

where

$$t_u^{mv}(2d-i\alpha_q)=\sum_{s=|u-v|}^{u+v}w_s h_s(k(2d-i\alpha_q))i^{u+v-p}, \tag{24}$$

and w_s is computed from the product of Wigner $3j$ symbols

$$w_s=\begin{pmatrix}v & u & s\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}v & u & s\\ m & -m & 0\end{pmatrix}\sqrt{(2u+1)\frac{(u-m)!}{(u+m)!}}\times\sqrt{(2v+1)\frac{(v-m)!}{(v+m)!}}. \tag{25}$$

The expressions given by Eq. (23) and Eq. (19) give a very compact and efficient way of computing the wave number integrals for the conversion and propagation of the spherical harmonics. It is interesting to note that in some sense the representation of the conversion integrals in terms of spherical harmonic translations (complex translations) is a generalization of the approach taken in the derivation of the analytic solution for a sphere near a perfectly reflecting surface, where the translations of spherical harmonics between the sphere and its perfect image are used.[13–15] In these references, a reflected version of the sphere (and the sphere, itself, may be rigid, a bubble, shelled, etc.) above the surface is considered in addition to the sphere in the water column (i.e., a negative version of the sphere placed at the standard image point). The system of equations for these coupled spheres is then derived and solved. The derivation of the system of equations for the two spheres involves representing the spherical harmonics of the image sphere's coordinate system in terms of the water column sphere's system. Thus, the expressions from the vertical translation of spherical coordinate systems naturally arise in this formulation. For a penetrable water/seabed boundary, we have used the method of complex images to facilitate the evaluation of various wave number integrals; however, for a sphere, one could instead start by considering it and its complex images (i.e., images placed at the complex positions determined by the α_q) with the corresponding amplitude term, and derive the system of simultaneous equations for this system in a manner analogous to the scenario with a perfectly reflecting flat surface. Because of this interpretation it is not surprising that the expression of Eq. (23) involves the terms which arise from the vertical translation of the spherical coordinate sys-

tems. However, for a penetrable boundary, these translations are complex, corresponding to the complex values of $i\alpha_q$.

In this paper we are not concerned directly with the numerical determination of the complex parameters (a_q, α_q), $q=1,...,Q$, but, of course, it is a necessary part of the overall complex-image method proposed. It is a nonlinear estimation problem, so that the potential for nonunique minima, nonconvergence of algorithms, etc., certainly exists. In Refs. 6–11 various numerical methods for determining the complex parameters are discussed. For all the various exponential fits of this paper, we used a very simple iterative nonlinear least-squares algorithm with damping (a simple form of the Levenberg–Marquardt algorithm[18]). The initial values of the parameters were simply taken to be $a_q \equiv 1$, and α_q set to be a linear distribution of positive imaginary values. For the case of a target on a homogeneous seabed, the estimation of the complex parameters was robust with a wide range of initial parameters converging to values which gave a very good fit. For some of the more difficult estimation cases (for example, the layered seabed, some of the transmission integrals discussed below) two or three different distributions of α_q had to be tried and the iterative algorithm damping factor varied before good estimates of the parameters a_q, α_q were obtained. Another approach which was used, in example for Ref. 8 for a waveguide, where a large number of images was required, was to simply fix the values of α_q and compute the values of a_q using a linear least-squares algorithm. In the future, it would be desirable to determine a more automated algorithm for the optimal determination of these complex parameters. In general, one would wish to minimize the number of coefficients, and this is straightforward in the case of a simple seabed, using the nonlinear least-square approach.

When the seabed is homogeneous, it can be shown[8] that the image parameters for $R(h)$ only need to be computed for a single frequency f_0 and then at any other frequency f

$$a_q(f)=a_q(f_0), \quad \alpha_q(f)=(f_0/f)\alpha_q(f_0). \tag{26}$$

This means that in multifrequency computations, it is easy to loop over the expressions given by Eq. (21), Eq. (23), and Eq. (24). This result is due to the fact that if one rewrites the horizontal wave number h as $h=\omega\tilde{h}$, then Eq. (17) can be written in a frequency-independent form

$$R(\tilde{h})=\frac{\rho_2\hat{\gamma}_1(\tilde{h})-\rho_1\hat{\gamma}_2(\tilde{h})}{\rho_2\hat{\gamma}_1(\tilde{h})+\rho_1\hat{\gamma}_2(\tilde{h})}, \tag{27}$$

where

$$\hat{\gamma}_i \equiv \sqrt{1/c_i^2-\tilde{h}^2}. \tag{28}$$

In the case that there is attenuation in the seabed, then the form of Eq. (27) can still be used if the attenuation in the seabed can be modeled by simply adding a negative imaginary constant (frequency independent) to the sound speed c_2. This is an approximation often used in acoustic modeling and is equivalent to attenuation varying linearly with frequency or specifying attenuation in the units dB/λ. The formula for the imaginary part of c_2 can be determined from, for example, Ref. 19 [Eq. (2.185), p.134]. Of course, this is a simplifying assumption; it relies upon the simple form of the reflection coefficient for a fluid/fluid interface and also assumes that attenuation varies linearly with frequency over a fixed bandwidth. For general viscoelastic media,[20] the frequency dependence of the reflection coefficient is more complicated, and for these cases one would have to perform a complex image fit at each frequency rather than using Eq. (26).

The discussions above were for the case of a target above (or on) the seabed. If the target is buried, the theory is almost identical for the conversion coefficients, Eq. (9), and the T-matrix system, Eq. (10). There is, however, a significant difference when one wishes to propagate the computed scattered field from the object up into the water column. In this case, the expression for the wave field in the water column due to the transmission of a particular spherical harmonic is given by Eq. (13). The complication in this case is that there are now two exponential factors with the different vertical wave numbers γ_w and γ_2 (the vertical wave numbers in water and sediment, respectively). In the case that the receivers are on the interface, $z=d$, the exponential term with the water vertical wave number is unity, and thus the integral becomes analogous to the reflection case; the complex parameters are chosen such that

$$\sum_{q=1}^{Q} a_q e^{\gamma_2 \alpha_q} \approx \Gamma(h). \tag{29}$$

For $z \neq d$, a fit for $\Gamma(h)\exp(i\gamma(z-d))$ could also be determined, but this would then involve a different fit for each value of z.

Another approach is to consider the expression for the mth azimuthal term of the scattered field in the water column, which results from summing Eq. (13) over the scattering coefficients

$$p_m(r,z)=i^{-m}\int_0^\infty J_m(hr)\left[e^{i\gamma_2 d}\Gamma(h)\sum_{n=m}^{N} p_{nm}^{sc} i^n P_n^m\left(\frac{\gamma_2}{k_2}\right)\right] \times \frac{e^{i\gamma_w(z-d)}}{k_2\gamma_2} h\,dh, \tag{30}$$

where p_{nm}^{sc} denotes the coefficients computed from solving the T-matrix equations. This expression can be rewritten in the form

$$p_m(r,z)=i^{-m}\int_0^\infty J_m(hr)F_m(h)\frac{e^{i\gamma_w(z-d)}}{k_w\gamma_w} h\,dh, \tag{31}$$

where

$$F_m(h) \equiv \frac{k_w\gamma_w(h)}{k_2\gamma_2(h)}\Gamma(h)\exp(i\gamma_2(h)d)\sum_{n=m}^{N} i^n p_{nm}^{sc} P_n^m\left(\frac{\gamma_2}{k_2}\right). \tag{32}$$

We now fit a sum of exponentials with complex parameters to $F_m(h)$ with respect to the vertical wave number in the water, γ_w. The term $\exp(i\gamma_2(h)d)$ is oscillatory as a function

of γ_2 (for real γ_2); however, d is often relatively small (shallow burial) for many cases of interest. Let us now consider Eq. (31) for a single exponential term (in practice we have to perform a sum of Q of these terms)

$$p_m^{q_m}(r,z) = i^{-m} a_{q_m} \int_0^\infty J_m(hr) e^{\gamma_w \alpha_{q_m}} \frac{e^{i\gamma_w(z-d)}}{k_w \gamma_w} h\, dh. \tag{33}$$

In the case of $m=0$ this integral is simply $a_{q_0} h_0(kR_{q_0})$, where $R_{q_0} \equiv \sqrt{r^2+(z-d-i\alpha_{q_0})^2}$. For $m \neq 0$ the integral is not so obvious. However, one can include the factor h of Eq. (33) with $F_m(h)$ and the resulting exponential fit, i.e.,

$$hF_m(h) = \sum_{q_m=1}^{Q} a_{q_m} \exp(\alpha_{q_m} \gamma(h)). \tag{34}$$

Substituting Eq. (34) for $hF_m(h)$ in the integrand of Eq. (30), we can then use the formulas from Ref. 21 [6.737 (1 and 2)] (with an appropriate definition of variables) or the formulas from Ref. 22 [8.7 (22) and 8.7 (47)], for real values of z and k_w, to write

$$\int_0^\infty J_m(hr) e^{\gamma_w \alpha_q} \frac{e^{i\gamma_w(z-d)}}{k_w \gamma_w} dh = \frac{\pi}{2k_w} J_{m/2}\left(\frac{k_w}{2}(R_{q_m} - \overline{z})\right) H_{m/2}^1\left(\frac{k_w}{2}(\overline{z} + R_{q_m})\right), \tag{35}$$

where $\overline{z} \equiv (z-d) - i\alpha_q$ and $H_{m/2}^1$ denotes a cylindrical Hankel function of the first kind. We numerically verified the validity of Eq. (35) for our subsequent computations using a variety of complex values of z and k_w. This equation is then used for each term of Eq. (34). The expression of Eq. (35) is no longer that for a simple point source, but it has a simple analytic form.

In summary, for this approach we compute the functions $F_0(h)$ and $hF_m(h)$ for azimuthal orders $m>0$ and perform an exponential fit to them to obtain the parameters (a_{q_m}, α_{q_m}). The function $F_m(h)$ is basically equivalent to the scattering solution at the mth azimuthal order, in wave number space, evaluated along the interface. From these exponential fits, we can evaluate the mth-order transmitted field in the water column from the relations

$$p_0(r,z) = \sum_{q_0=1}^{Q} a_{q_0} h_0(kR_{q_0}), \tag{36}$$

for $m=0$ and

$$p_m(r,z) = i^{-m} \frac{\pi}{2k_w} \sum_{q_m=1}^{Q} a_{q_m} J_{m/2}\left(\frac{k_w}{2}(R_{q_m} - \overline{z})\right) \times H_{m/2}^1\left(\frac{k_w}{2}(\overline{z} + R_{q_m})\right), \tag{37}$$

for $m>0$.

For odd values of m in Eq. (37), the cylindrical functions needed become spherical Bessel and Hankel functions, as the spherical functions are simply related to the odd half-order cylindrical functions. For numerical purposes, one might wish to avoid the computation of cylindrical functions altogether, because spherical functions are easier to compute. It is possible to do so by finding a way to evaluate Eq. (33) for at least the even values of $m>0$. This may be done as follows. Suppose the expansion coefficients (a_{q_M}, α_{q_M}) are known for a given $F_M(h)$ where M is even. For $m \leqslant M$, define the series of integrals

$$p_m^{q_M}(r,z) \equiv i^{-M} a_{q_M} \int_0^\infty J_m(hr) e^{\gamma_w \alpha_{q_M}} \frac{e^{i\gamma_w(z-d)}}{k_w \gamma_w} h\, dh, \tag{38}$$

for m even and

$$p_m^{q_M}(r,z) \equiv i^{-M} a_{q_M} \int_0^\infty J_m(hr) e^{\gamma_w \alpha_{q_M}} \frac{e^{i\gamma_w(z-d)}}{k_w \gamma_w} dh, \tag{39}$$

for m odd. The odd m integrals are, of course, known in terms of spherical Bessel functions through Eq. (35). For the even m integrals, we substitute the recurrence relation[19]

$$J_{m+1}(x) = \frac{2m}{x} J_m(x) - J_{m-1}(x), \tag{40}$$

into Eq. (38) so that when $m \pm 1$ is even we obtain

$$p_{m+1}^{q_M}(r,z) = \frac{2m}{r} p_m^{q_M}(r,z) - p_{m-1}^{q_M}(r,z). \tag{41}$$

Recalling that

$$p_0^{q_M}(r,z) = a_{q_M} h_0(kR_{q_M}), \tag{42}$$

and using Eq. (35) for the odd values of m, it is then possible to compute each term needed in $p_M(r,z)$ without explicitly computing cylindrical Bessel functions through recursion of Eq. (41).

We have outlined a complex-image (generalized images) approach for the transmission problem. In the numerical examples we will see that this approach is efficient and very accurate. However, it is clear that this case is still more complicated than for the reflection scenario. It is necessary to perform a fit for each azimuthal order. Additionally, the complex image parameters for the transmission case do not have the simple frequency dependence that the reflection images have for a homogeneous seabed.

III. NUMERICAL EXAMPLES

In the first numerical example we consider a seabed with $c_p = 1700$ m/s and an attenuation of 0.5 dB/λ and $\rho = 1.5$ g/cm^3. The water has sound speed $c_p = 1500$ m/s and $\rho = 1.0$ g/cm^3. The attenuation is modeled by adding an appropriate imaginary part to the sound speed. As discussed previously, for this simple type of fluid/fluid interface, the expressions of Eq. (26) can be used to efficiently compute the complex image parameters for multiple frequencies. In Fig. 2 we show the true reflection coefficient (computed at 1500 Hz) and the approximate reflection coefficient using a six-term approximation (with one of the terms being the asymptotic term) as a function of the normalized horizontal wave number (i.e., frequency factored out). As can be seen, the method of complex images has provided an excellent fit. The conver-

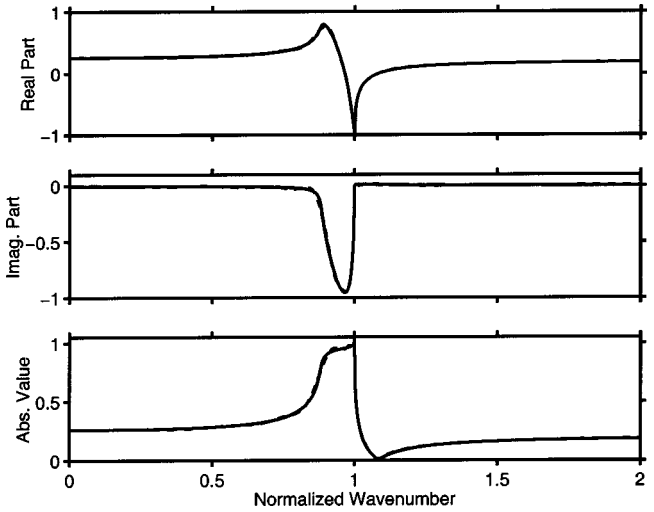


FIG. 2. The exact (solid line) and complex-image (six terms including asymptotic term) (dashed line) reflection coefficient for the homogeneous seabed: $c_p = 1700$ m/s, $\rho = 1.5$ g/cm^3, and attenuation, $\mu = 0.5$ dB/λ. A comparison of the real parts, imaginary parts, and the absolute values is shown in the three panels.

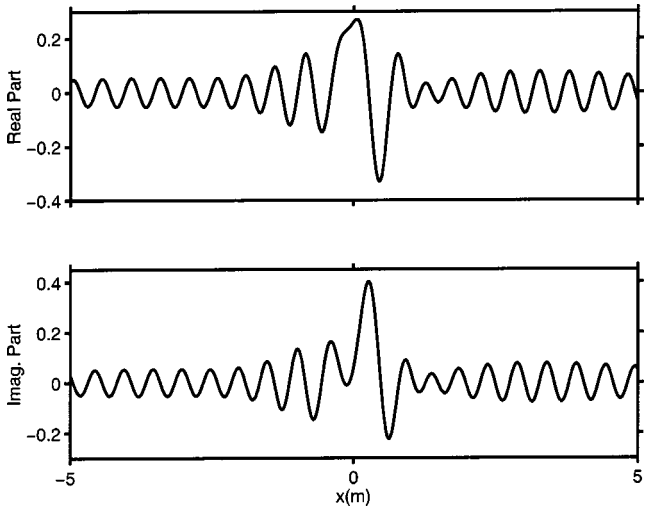


FIG. 4. The scattered field at 3 kHz for 0.25-m radius rigid sphere sitting on the homogeneous seabed using the exact integrals (solid line) and the complex-image approximations (dashed line). The receivers are along $z = 0.5$ m (i.e., 0.25-m above the top of the sphere).

sion coefficient $C^1_{2,1}$ is computed using the exact integral and the complex-image coefficient approximation, Eq. (23). The comparison between the real and imaginary parts for the two approaches as a function of frequency is shown in Fig. 3. As a function of frequency, it can be seen from Eq. (23) that the magnitude of these coefficients will grow very large for small frequency because of the behavior of $h_n(kd)$ for small argument. This is evident in Fig. 3, for the imaginary part, where it can be seen that the magnitude of the coefficient is becoming larger at the lower frequencies. It can be seen that there is excellent agreement between the complex image method and the integral. (There is a small disagreement for the real parts at low frequencies, but the real part is insignificant in comparison to the imaginary part, for which the two methods have excellent agreement.)

We now compute the scattered pressure field along a line of receivers using the ''exact'' numerical integration for the conversion matrix and the reflection of the scattered field. The same computation is then performed using the expressions derived with the complex-image method. We do this for a rigid sphere of radius 0.25 m sitting on the seabed (with parameters as above) and a frequency of 3 kHz. The line of receivers is located at 0.5 m (0.25 m above the top of the sphere) for $y = 0$ and x varying between -5 to 5 m. Negative values of x represent the backscatter direction and positive values the forward direction. The incident field is a plane wave incident at an angle of 20 deg from the vertical. We also do the same computations for x varying between 50 and 60 m (100 to 120 wavelengths at 3 kHz). The results displayed in Figs. 4 and 5 show that the scattered fields computed by using the expressions provided by the method of

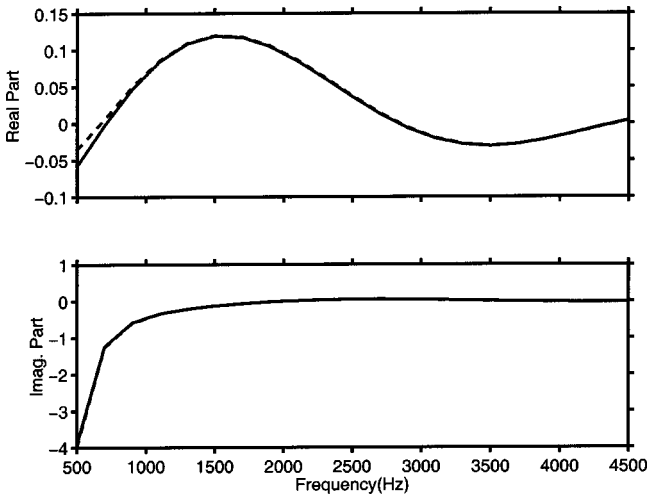


FIG. 3. The real and imaginary parts of the conversion coefficient for the $(n=1, m=1)$ harmonic into the $(n=2, m=1)$ harmonic as computed using the exact integral (solid line) and complex-image approximation (dashed line) as a function of frequency for the homogeneous seabed of Fig. 2.

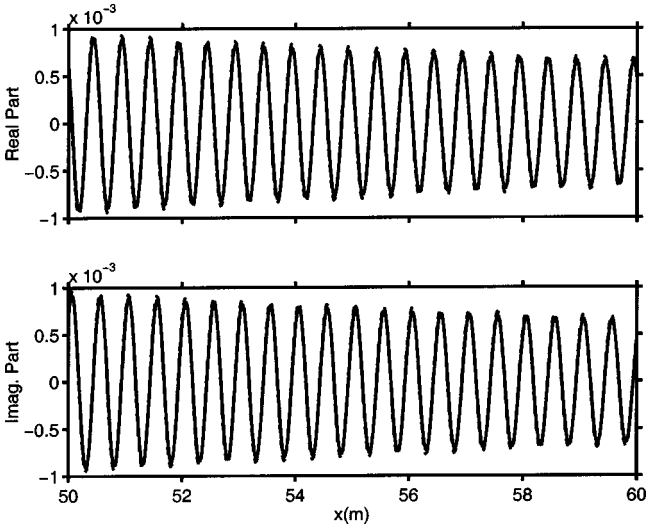


FIG. 5. The scattered field at 3 kHz for a 0.25-m radius rigid sphere sitting on the homogeneous seabed using the exact integrals (solid line) and the complex-image approximations (dashed line). The horizontal range is from 50 to 60 m (or 100 to 120 wavelengths). The receivers are along $z = 0.5$ m (i.e., 0.25 m above the top of the sphere).

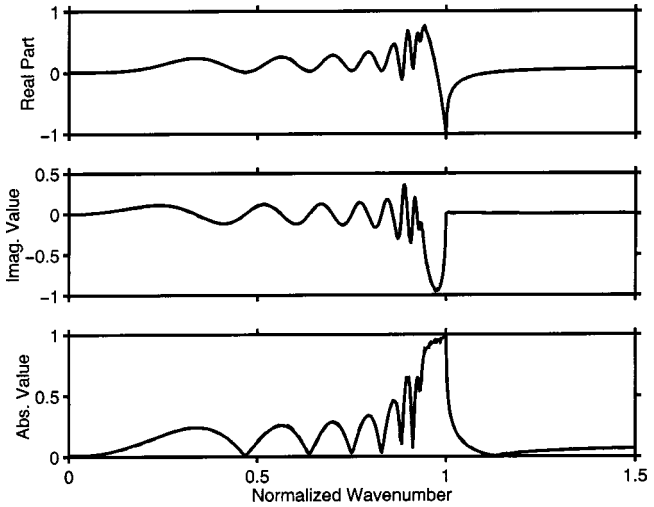


FIG. 6. Comparison of exact reflection coefficient (solid line) and complex image fit (14 terms including asymptotic term) (dashed line) for a layered seabed at 3 kHz. There is a 2-m-thick sediment layer ($c_p = 1600$ m/s, $\rho = 1.2$ g/cm^3 overlying the half-space ($c_p = 1700$ m/s, $\rho = 1.5$ g/cm^3. The attenuation is 0.25 dB/λ throughout the seabed. A comparison of the real parts, imaginary parts, and the absolute values is shown in the three panels.

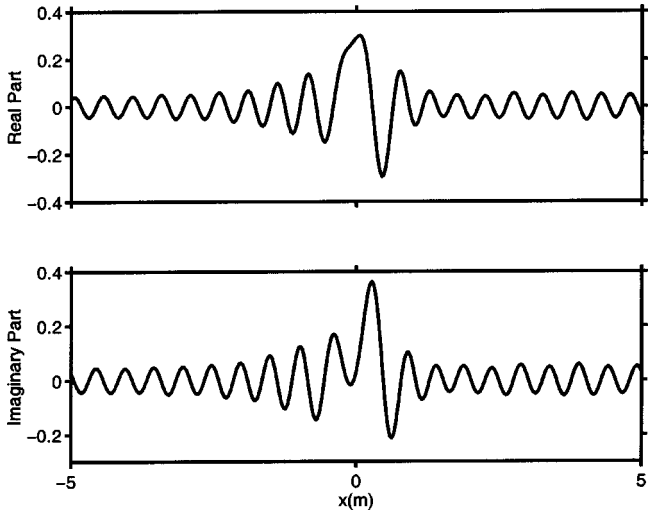


FIG. 7. The scattered field at 3 kHz for a 0.25-m radius rigid sphere sitting on the layered seabed using the exact integrals (solid line) and the complex-image approximations (dashed line). The receivers are along $z = 0.5$ m (i.e., 0.25 m above the top of the sphere). The angle of incidence is 20° off the vertical.

complex images are in very good agreement with the benchmark solutions obtained by directly computing all the wave number integrals.

Let us now consider a layered seabed with a 2-m-thick sediment layer ($c = 1600$ m/s and $\rho = 1.2$ g/cm^3) overlying a half-space ($c = 1700$ m/s and $\rho = 1.5$ g/cm^3). The attenuation is 0.25 dB/λ for the entire seabed. The frequency is 3 kHz. As can be seen in Fig. 6, the reflection coefficient is much more complicated as a function of horizontal wave number than for a homogeneous half-space, and hence a larger number of terms is required for an accurate fit. The complex-image fit to the reflection coefficient using 14 exponentials is also shown in Fig. 6. The parameters from this exponential fit are then used as before in the expressions to set up the scattering matrix and to propagate the scattered field; a comparison of the scattered pressure field using the complex-image evaluation of all the integrals and the computations using the direct evaluation of all the wave number integrals is shown, for a line of receivers 0.25 m above the top of the sphere, in Fig. 7. The angle of incidence is 20° off the vertical. As can be seen, the agreement of the solution obtained using the method of complex images with the benchmark solution is almost exact.

It should be emphasized that the results obtained thus far using the complex-image expressions involved no wave number integrals, only the expressions given by Eqs. (21) and (23). The results of this paper show the accuracy of these expressions for computing the scattered fields. The relative numerical efficiency of using these sums depends on how efficiently one can compute these small sums in comparison to using the sums resulting from the direct discretization of the wave number integrals. It is not possible to give a definitive conclusion regarding the efficiency of the complex-image expressions as that is related to many factors: the geometry of the receivers, the programming language used, the efficiency of computing the special functions, etc. In the implementations of this paper, we used the programming language MATLAB.[23] Numerically, we used the trapezoidal rule to evaluate the wave number integrals (the appropriate discretization and the length of the integral was found by prior numerical experimentation). The direct implementation of the wave number integrals was made very efficient by minimizing the amount of repeat computations and utilizing vector commands wherever possible. It was found that the complex-image expression for the propagation of the scattered field, Eq. (21), provided significant computational savings over the direct implementation of the corresponding wave number integrals. For example, for the homogeneous seabed and the ranges of 50–60 m, it was approximately 15 times faster than directly using the wave number integrals. For the conversion integrals, the direct implementation of the wave number expressions and the complex-image expression had essentially equal computation times. Even though the expression, Eq. (23), involves only Q terms, the computation of the translation coefficients t_u^{mv}, including the computation of the Wigner $3j$ symbols, is quite time consuming (in a relative sense) in our implementation. Certainly, for multifrequency computations, we would compute many of these quantities only once, and thus improve the overall efficiency of this part of the computation.

We now consider an example with a buried sphere at a frequency of 3 kHz. The sphere is buried 50 cm below the seabed interface (25 cm from the top of the sphere). The seabed has the parameters $c_p = 1700$ m/s, $\rho = 1.5$ g/cm^3, and an attenuation of 0.5 dB/λ. There is an incident plane wave corresponding to an angle of incidence of 20° off the vertical in the water column. The reflection coefficient is now in terms of the reflection of a wave in the lower half-space off the sediment/water interface. This is simply the negative of the coefficient going from the opposite direction. However, there are significant differences between the two cases when considering the complex image fit. For the reflection coefficient in the water column, the complex image coefficients

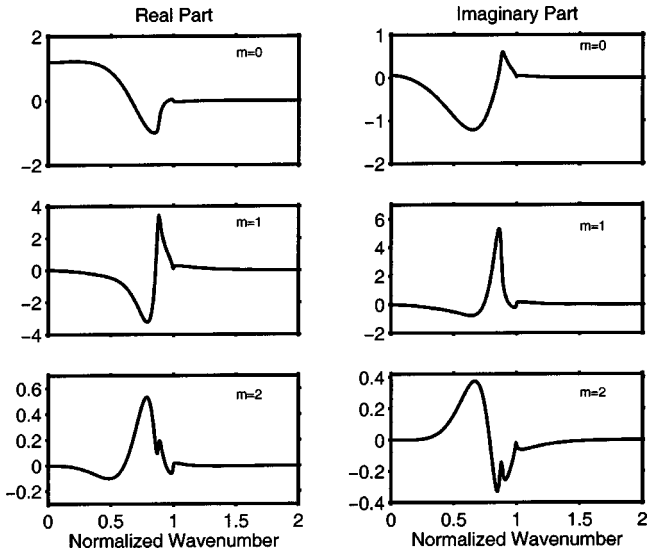


FIG. 8. Comparison between the exact values (real and imaginary parts) of $F_m(h)$, for $m=0$ and $hF_m(h)$, for $m=1,2$ (solid) and the complex-image fits (ten terms) (dashed) for the first three azimuthal orders. The frequency is 3 kHz and the sphere is buried 0.5 m below the interface.

were determined in terms of γ_w for the water. Now, the fit is in terms of γ_2

$$\sum_{q=1}^{Q} a_q \exp(\alpha_q \gamma_2) \approx R(h), \tag{43}$$

and thus the determined complex parameters are different than simply reversing the sign of the amplitude coefficients for the fit of reflection coefficient in the water. In fact, it was found that the fit in terms of γ_2 was slightly more difficult. A total of ten coefficients, nine determined coefficients and the tenth is the asymptotic term, are used in the following computations. In Fig. 8, the real and imaginary parts of $F_0(h)$ and $hF_m(h)$ [Eq. (32)] and the corresponding ten-term fits are shown for $m=0,1,2$ (we actually computed up to and including $m=6$ in the field computations). As can be seen,

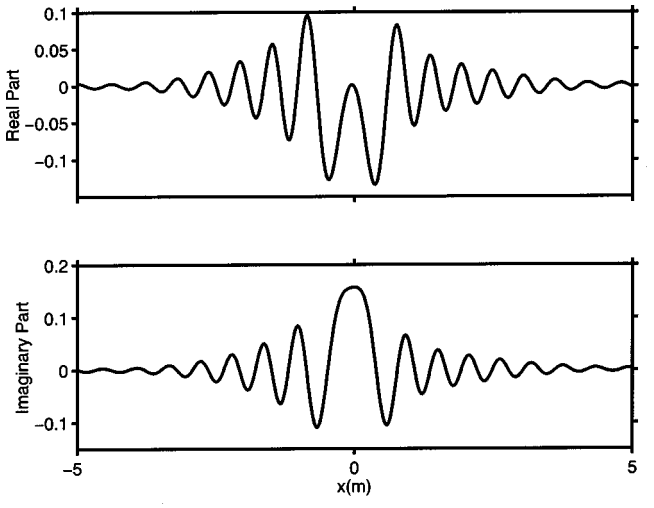


FIG. 9. Comparison between computed pressure field in the water column using integral expressions (solid) for T-matrix and propagation (transmission) and using complex images (dashed) for T-matrix and generalized images for propagation (transmission) computations for line of receivers at 0.25 m above the interface. The frequency is 3 kHz.

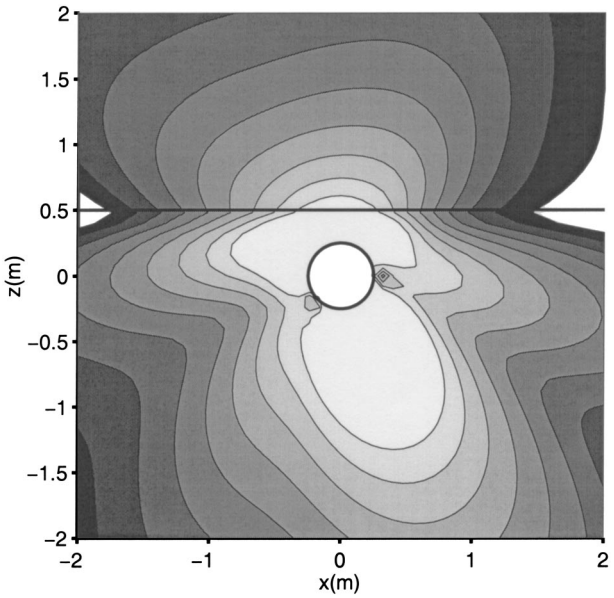


FIG. 10. Two-dimensional slice of field scattered by a buried sphere at 3 kHz for the homogeneous seabed: $c_p=1700$ m/s, $\rho=1.5$ g/cm^3, and $\mu=0.5$ dB/λ:levels are in dB from -30 to -14 dB in steps of 2 dB. All computations are done using complex image formulas. The angle of incidence in the water column is 20° off the vertical.

the fits are virtually identical to the true values of $F_0(h)$ and $hF_m(h)$. For this example, the function $F(h)$ did not become significantly more complicated as a function of h for increasing azimuthal order, and ten terms were quite sufficient for the approximation at each order. If the frequency of the problem was increased, one would expect $F(h)$ to become a more complicated function of h due to the fact that the solution to the scattering problem from the sphere will, in general, be more oscillatory as the frequency increases: also, there is the term $\exp(i\gamma(h)d)$ in Eq. (32) which becomes increasingly oscillatory as the maximum real value of γ increases with frequency. Thus, we would expect to have to use more complex-image terms for higher frequencies.

The exact and complex-image approximations to the transmitted field (0.25 m above the interface) are shown in Fig. 9 for a line of receivers in the water column. For the exact solution, the wave number integrals were used for the conversion integrals for the T-matrix and to propagate the scattered field; for the image method, images were used both to compute the T-matrix (using the reflection images) and the generalized images were used to compute the transmitted scattered field. As can be seen the complex-image approximation is almost indistinguishable from the exact computation. Finally, for Fig. 10 we use the complex-image expressions for all parts of the computation: T-matrix, and propagation of the scattered field, both reflected within the seabed and transmitted into the water to produce a two-dimensional slice ($y=0$) of the scattered field.

IV. SUMMARY

In this paper we have shown how the method of complex images can be used to obtain simple expressions for the wave number integrals of target/seabed scattering. Numerical examples established the accuracy of these expressions for

both the individual conversion coefficients and for computing the complete scattered field. The method was applied to both a homogeneous and layered seabed. For the case of a homogeneous seabed, the complex-image fit to the reflection coefficient need only be done at a single frequency in multifrequency computations, and hence the complex-image method is particularly powerful in this case.

The complex-image method allowed us to very efficiently compute the scattered field from a sphere on a seabed. For a buried target, the computation of the conversion and propagation within the seabed proceeds as for the target above the seabed. The computation of the scattered field transmitted back into the water is more problematic. For these computations, we derived a generalized image method which allowed us to replace the wave number integral computations with small sets of image terms for each azimuthal order.

It is difficult to definitively quantify the improvement in efficiency of using the expressions obtained from the complex image approach over the direct numerical quadrature of the wave number integrals for the computation of the T-matrix elements and the propagation of the spherical harmonics. For the implementations of this paper, using the expressions obtained from the method of complex images resulted in a significant improvement in the computational speed for the propagation of the scattered field. The complex-image expressions for the elements of the T-matrix (conversion integrals) were not as numerically efficient due to the time spent computing the individual terms in the sum, which are related to the vertical translation of spherical harmonics. However, it is anticipated that this type of computation (for example, by computing the Wigner $3j$ symbols only once in multifrequency computations) could be made more efficient. In addition, for the method of complex images there is a small overhead associated with the initial determination of the complex image parameters. As discussed in the text, we have not optimized or automated this procedure. It would be desirable to do this in the future.

ACKNOWLEDGMENTS

R. Lim gratefully acknowledges support for this work from ONR and SERDP.

[1] P. C. Waterman, "New formulation of acoustic scattering," J. Acoust. Soc. Am. **45**, 1417–1429 (1969).

[2] R. Lim, J. Lopes, R. Hackman, and D. Todoroff, "Scattering by objects buried in underwater sediments: Theory and experiment," J. Acoust. Soc. Am. **93**, 1762–1783 (1993).

[3] R. Lim, "Multiple scattering by many bounded obstacles in a multilayered acoustic medium," J. Acoust. Soc. Am. **92**, 1593–1612 (1992).

[4] G. C. Bishop and J. Smith, "Scattering from rigid and soft targets near a planar boundary: Numerical results," J. Acoust. Soc. Am. **105**, 130–143 (1999).

[5] J. Fawcett, W. L. J. Fox, and A. Maguer, "Modeling of scattering by objects on the seabed," J. Acoust. Soc. Am. **104**, 3296–3304 (1998).

[6] Y. L. Li and M. J. White, "Near-field computation for sound propagation above ground—Using complex image theory," J. Acoust. Soc. Am. **99**, 755–760 (1996).

[7] J. Fawcett, "Complex-image approximations to the half-space acousto-elastic Green's function," J. Acoust. Soc. Am. **108**, 2791–2795 (2000).

[8] J. Fawcett, "A method of images for a penetrable acoustic waveguide," J. Acoust. Soc. Am. **113**, 194–204 (2003).

[9] R. M. Shubair and Y. L. Chow, "A simple and accurate complex image interpretation of vertical antennas present in contiguous dielectric half spaces," IEEE Trans. Antennas Propag. **41**, 806–812 (1993).

[10] J. J. Yang, Y. L. Chow, and D. G. Fang, "Discrete complex images of a three-dimensional dipole above and within a lossy ground," IEE Proc., Part H: Microwaves, Antennas Propag. **138**, 319–326 (1991).

[11] N. Geng and L. Carin, "Wide-band electromagnetic scattering from a dielectric BOR buried in a layered lossy dispersive medium," IEEE Trans. Antennas Propag. **47**, 610–619 (1999).

[12] R. Hickling, "Analysis of echoes from a hollow metallic sphere in water," J. Acoust. Soc. Am. **36**, 1124–1137 (1964).

[13] G. C. Gaunaurd and H. Huang, "Acoustic scattering by a spherical body near a plane boundary," J. Acoust. Soc. Am. **96**, 2526–2536 (1994).

[14] G. C. Gaunaurd and H. Huang, "Acoustic scattering by an air-bubble near the sea surface," IEEE J. Ocean. Eng. **20**, 285–292 (1995).

[15] H. Huang and G. C. Gaunaurd, "Scattering of a plane acoustic wave by a spherical elastic shell near a free surface," Int. J. Solids Struct. **34**, 591–602 (1997).

[16] L. B. Felsen, "Complex-source-point solutions of the field equations and their relation to the propagation and scattering of Gaussian beams," Symp. Math. Inst. Mat. **18**, 39–56 (1976).

[17] G. Frisk, *Ocean and Seabed Acoustics: A Theory of Wave Propagation* (PTR Prentice-Hall, New York, 1994).

[18] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1988).

[19] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, Woodbury, NY, 1994).

[20] R. D. Borcherdt, "Reflection–refraction of general P- and type-I S-waves in elastic and anelastic solids," Geophys. J. R. Astron. Soc. **70**, 621–638 (1982).

[21] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed. (Academic, New York, 1965).

[22] *Tables of Integral Transforms*, Bateman Manuscript project, edited by A. Erdéyli (McGraw-Hill, New York, 1954), Vol. 2.

[23] *Using* MATLAB: *Version 6*, The Math Works, Inc., Natick, MA (2000).

Diffraction and attenuation of a tone burst in mono-relaxing media

Yefim Dain and Richard M. Lueptow[a)]
Department of Mechanical Engineering, 2145 Sheridan Road, Northwestern University, Evanston, Ilinois 60208

(Received 6 August 2002; revised 20 June 2003; accepted 30 June 2003)

The influence of intrinsic absorption in a relaxing medium and the resulting three-dimensional diffraction correction of the magnitude of the acoustic pressure averaged over the surface of a receiver is investigated for a tone burst. A rigorous formula for the damped acoustic pressure average at the receiver was obtained for arbitrary pulsed waves in a mono-relaxing medium. Depending on the pulse oscillation frequency, envelope duration, and relaxation frequency of the media, the plane wave burst envelope can be reduced, amplified, or otherwise deformed. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1602701]

PACS numbers: 43.35.Ae, 43.35.Fj, 43.20.El [JGH]

I. INTRODUCTION

The prediction of diffraction for a transient burst propagating through an acoustic medium that exhibits relaxation is of interest for interpreting measurements of acoustic attenuation. In this paper, we develop a theory for the three-dimensional diffraction of a tone burst signal in a relaxing medium. Previous research has addressed separately various pieces of the problem such as attenuation due to three-dimensional diffraction of a harmonic wave, attenuation due to relaxational losses of a plane harmonic wave, or attenuation of a single plane pulse due to relaxational losses. However, to our knowledge no previous study has addressed the combined problem of attenuation due to diffraction (for a three-dimensional wave rather than for plane wave) for a burst signal (as opposed to a harmonic wave or a single pulse), while including attenuation due to relaxational losses (as opposed to ignoring relaxational losses). It is this problem that is addressed here.

We assume that a rigid circular planar piston with a uniform distribution of normal velocity on its surface is embedded in a rigid infinite baffle and radiates into the half-space of an ideal homogeneous, isotropic, relaxing medium. An acoustically transparent receiver with uniform sensitivity over its surface and a uniform amplitude-frequency response is coaxial with the piston. We are interested in the rigorous calculation of the average pressure exerted on the receiver by the three-dimensional acoustic pressure burst after propagating through a relaxing medium. Although a comprehensive investigation of this problem has not been reported, certain aspects of the problem have been considered by previous researchers.

Several methods to correct for the attenuation of a three-dimensional harmonic steady-state pressure wave due to diffraction (but not attenuation due to relaxational losses) have been described based on various models of the harmonically excited field in a loss-less medium.[1] Seki *et al.*[2] used approximate formulas for the average pressure over the piston in the absence of absorption based on the Lommel[3] and Williams[4] harmonic wave solutions of the Helmholtz equation in an attempt to account for diffraction effects in the measurement of acoustic attenuation. Khimunin[5,6] also applied the loss-less formula of Williams[4] for the average harmonic pressure to the diffraction correction of attenuation, demonstrating the influence of intrinsic absorption on the harmonic average pressure of the piston. Rogers and VanBuren[7] obtained a simple expression for the diffraction correction using Lommel's expression, which is a high-frequency limit of Williams' exact expression.

In addition to attenuation due to diffraction, classical attenuation and relaxational attenuation reduce the amplitude of a harmonic acoustic wave as it propagates. The theory of harmonic wave attenuation in relaxing medium was formulated by Kneser[8] and Mandel'shtam and Leontovich.[9] There is a substantial experimental and theoretical literature addressing relaxation in gases. We have recently extended the theoretical prediction of relaxational attenuation for a harmonic wave to a three-component gas mixture.[10]

The evolution of a single plane pulse (rather than a harmonic wave) in the absence of diffraction or scattering in relaxing medium has also been studied. A thorough analysis of the propagation of plane pulses in a relaxing media with relaxation laws modeling a variety of homogeneous materials was made in the review of Vainshtein,[11] with a focus on electromagnetic waves. Detailed analyses of the propagation of a single plane-wave acoustic pulse in the case of a mono-relaxing medium were performed analytically and numerically by Dunin,[12] Dunin and Maksimov,[13] Andreev *et al.*,[14] Larichev and Maksimov,[15] and Andreev *et al.*[16] Their results show a broadening of the pulse and a decrease in the pulse amplitude as the pulse propagates in space. In some cases, the theory was shown to match experiments reasonably well,[14,16] although in other cases differences between the measurement and theory were attributed to diffraction ef-

a)Author to whom correspondence should be addressed. Electronic mail: r-lueptow@northwestern.edu

fects, which are not taken into account when assuming a plane wave.[16]

The aim of this paper is to calculate analytically the diffraction correction to the attenuation of a three-dimensional tone burst in a mono-relaxing gas medium. To put this into context, the three-dimensional effects of diffraction on a harmonic wave in a lossless medium have been studied, as have the effects of relaxational losses on a harmonic wave or a single pulse for a one-dimensional plane wave. However, in order to recover the attenuation of sound from practical experiments in a relaxing medium, it is essential to formulate a model that can describe both the attenuation due to relaxational acoustical properties of the relaxing medium and the attenuation due to the three-dimensional distortion of acoustic wave. In addition, since many experiments use a short burst of sound to avoid reflections in the measured quantities, neither the harmonic wave nor the single pulse analysis is appropriate. Thus, the aim of this paper is to calculate analytically the diffraction correction to the attenuation of a three-dimensional tone burst in a mono-relaxing gas medium.

The solution will be based on the application of a Green's function to the three-dimensional wave equation for acoustical pressure in a medium with relaxation and averaging the pressure over the surface of the receiver. Consequently, the dependence of total attenuation of the average acoustic pressure on diffraction and relaxational attenuation can be separated. A result is that the phenomenon of distortion of the complex envelope of the oscillating pulse as it propagates can be represented by a relation that depends on the relaxation time of medium and the frequency of pulse oscillation.

II. THEORETICAL TREATMENT

A linear acoustic wave equation for acoustic pressure in a mono-relaxing gas medium has the form suggested by Rudenko and Soluyan,[17]

$$\frac{\partial^2 p}{\partial t^2} - c_0^2 \Delta p - m c_0^2 \Delta \int_{-\infty}^{t} \frac{\partial p}{\partial t'} e^{-(t-t')/\tau_{\text{relax}}} \partial t' = 0, \tag{1}$$

where τ_{relax} is the relaxation time for the relaxation process, $m=(c_\infty^2-c_0^2)/c_0^2$ is the net increase in phase speed as frequency varies from zero to infinity which characterizes the relaxation strength, c_0 is the equilibrium speed of sound, and c_∞ is the frozen speed of sound. Typically, parameter m satisfies the condition $m \ll 1$.

We assume a transient source condition for normalized pressure,

$$p(r,0,t)=H(a-r)M(t)e^{i\omega_0 t}, \tag{2}$$

where H is the Heaviside function, r is radial coordinate, a is the radius of the piston, and the burst envelope $M(t)$ is a slowly varying function of time in comparison with the period of the oscillating component, $e^{i\omega_0 t}$.

The average pressure on an acoustically transparent receiver placed coaxial in relation to the piston source depends on the spacing z between transducer and receiver so that

$$\langle p(z,t)\rangle = \frac{1}{S}\int p(r,z,t)\,dS, \tag{3}$$

where S is the area of receiver and angled brackets denote an averaged value.

We suppose for brevity that the radius of the receiver is equal to that of the transducer. In order to calculate the average pressure of Eq. (3), the Fourier transform in the time domain is applied to Eq. (1) using boundary condition Eq. (2). Then the average pressure can be represented for all frequencies in the form of Fourier integral, where use was made of the formula of average harmonic pressure of Williams[4]

$$\begin{aligned}\langle p(z,t)\rangle &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\langle p(z,\omega)\rangle e^{i\omega t}\,d\omega \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\Bigg[e^{-ik(\omega)z} \\ &\quad - \frac{4}{\pi}\int_0^{\pi/2} e^{-ik(\omega)(z^2+4a^2\cos^2\theta)^{1/2}}\sin^2\theta\,d\theta\Bigg] \\ &\quad \times \overline{M}(\omega-\omega_0)e^{i\omega t}\,d\omega.\end{aligned} \tag{4}$$

The expression in the brackets is identical to that derived by Williams[4] for a harmonically oscillating piston in a lossless medium, where the wave number is real. In Eq. (4), the wave number $k(\omega)$, which is the dispersion relation of relaxing medium, is a complex function of frequency. The first term in the brackets is that for a plane harmonic wave; the second term in the brackets is the diffraction correction for a plane acoustic wave.

If the envelope function in Eq. (2) is set to unity, $M(t)=1$, the piston source becomes a harmonic oscillation. In such a case the Fourier transform of $M(t)$ is the Dirac delta-function, and the formula for the average oscillating pulse pressure (4) simplifies to the formula of the average harmonic pressure of Williams [his Eq. (16)],[4] which is based on Eq. (1) with the relaxation terms omitted. Therefore, Eq. (4) for the average pressure of a burst in a mono-relaxing medium is a generalization of the Williams formula (which is limited to a harmonic wave in a lossless medium).

The dispersion relation for Eq. (1) can be approximated following Vainshtein[11] and Andreev *et al.*[14] by the expression

$$k(\omega)=\frac{\omega}{c_0}\left(1+\frac{m\omega\tau_{\text{relax}}}{1+i\omega\tau_{\text{relax}}}\right)^{1/2} \approx \frac{\omega}{c_\infty}+\frac{m'\omega}{2c_\infty(1+i\omega\tau_{\text{relax}})} \tag{5}$$

with an accuracy of m'^2, where $m'=2(c_\infty-c_0)/c_0$. The two forms for the net increase in phase speed, m and m', are of the same order of magnitude because $m-m'=(c_\infty-c_0)^2/c_0^2 \ll 1$.

The dispersion relation Eq. (5) can be used to transform the average pressure Eq. (4) into a form suitable for the separate investigation of the attenuation of the oscillating portion of the burst and the envelope of the burst for the source Eq. (2). First, let

$$\xi=\omega-\omega_0 \tag{6}$$

so that the Fourier integral Eq. (4) becomes

$$\langle p(z,t)\rangle \approx \frac{e^{i\omega_0 t}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left[e^{-ik(\xi+\omega_0)z} - \frac{4}{\pi} \int_0^{\pi/2} e^{-ik(\xi+\omega_0)(z^2+4a^2\cos^2\theta)^{1/2}} \sin^2\theta\, d\theta \right] \times \overline{M}(\xi) e^{i\xi t}\, d\xi. \quad (7)$$

The form of the dispersion relation Eq. (5) makes possible the algebraic transformation in Eq. (7), leading to a modified dispersion relation of the form

$$k(\omega_0+\xi) = \mathrm{k}(\omega_0) + k_{\mathrm{mod}}(\xi), \quad (8)$$

where the modified dispersion parameter

$$k_{\mathrm{mod}}(\xi) = \frac{\xi}{c_\infty} + \frac{m''\xi}{2c_\infty(1+i\xi\tau''_{\mathrm{relax}})} \quad (9)$$

depends on new parameters of relaxation

$$m'' = \frac{m'}{[1+i\tau_{\mathrm{relax}}\omega_0]^2}, \quad \tau''_{\mathrm{relax}} = \frac{\tau_{\mathrm{relax}}}{[1+i\tau_{\mathrm{relax}}\omega_0]}. \quad (10)$$

Since the integrated function in Eq. (7) decays exponentially as ξ approaches infinity, the Fourier integral Eq. (7) can be represented using the dispersion relation Eq. (8) as

$$\langle p(z,t)\rangle = \frac{e^{i\omega_0 t - ik(\omega_0)z}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{i\xi t - ik_{\mathrm{mod}}(\xi)z} \overline{M}(\xi)\, d\xi - \frac{4e^{i\omega_0 t}}{\pi} \times \int_0^{\pi/2} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{i\xi t - ik_{\mathrm{mod}}(\xi)(z^2+4a^2\cos^2\theta)^{1/2}} \times \overline{M}(\xi)\, d\xi \right) e^{-ik(\omega_0)(z+4a^2\cos^2\theta)^{1/2}} \sin^2\theta\, d\theta. \quad (11)$$

The Fourier integrals on the right-hand side of Eq. (11) can be replaced by a convolution integral using the well-known expression for the Green's function as suggested by Vainshtein[11] and Dunin,[12]

$$g_{\mathrm{mod}}(\zeta'', t_{\mathrm{ret}}) = \exp(-\zeta'')\,\delta(t_{\mathrm{ret}}) + H(t_{\mathrm{ret}}) \times \left(\frac{\zeta''}{t_{\mathrm{ret}}\tau''_{\mathrm{relax}}} \right)^{1/2} \exp\left(-\left| \zeta'' + \frac{t_{\mathrm{ret}}}{\tau''_{\mathrm{relax}}} \right| \right) \times I_1[2(\zeta'' t_{\mathrm{ret}}/\tau''_{\mathrm{relax}})^{1/2}], \quad (12)$$

which describes the evolution of a delta-function input in a relaxing medium with modified parameters of relaxation Eq. (10). Here I_1 is the modified Bessel function of the first kind. In addition, t_{ret} is the retarded time and ζ'' is the dimensionless complex number depending on the separation z having the forms

$$t_{\mathrm{ret}} = t - z/c_\infty, \quad \zeta'' = m''z/2c_\infty\tau''_{\mathrm{relax}}. \quad (13)$$

Using variables Eq. (13) for brevity, the convolution integral Eq. (11) takes the form

$$\langle p(z,t)\rangle = e^{i\omega_0 t - ik(\omega_0)z} \int_{-\infty}^{+\infty} \left[g_{\mathrm{mod}}(\zeta'', t_{\mathrm{ret}} - s) - \frac{4e^{ik(\omega_0)z}}{\pi} \times \int_0^{\pi/2} g_{\mathrm{mod}}(Z''(\theta), t_{\mathrm{ret}} - s) \times e^{-ik(\omega_0)Z(\theta)} \sin^2\theta\, d\theta \right] M(s)\, ds, \quad (14)$$

where function $\zeta''(\theta)$ is calculated by means of Eq. (13), s is a dummy parameter of integration, and

$$Z(\theta) = (z^2+4a^2\cos^2\theta)^{1/2}, \quad Z''(\theta) = \left(\zeta''^2 + 4\left(\frac{2c_\infty\tau''_{\mathrm{relax}}a}{m''} \right)^2 \cos^2\theta \right)^{1/2}. \quad (15)$$

Equation (14) is a generalization of the well-known expression for the diffraction correction of harmonic wave in a lossless medium of Williams.[4] However, Eq. (14) permits a burst via the envelope $M(t)$ and includes relaxational effects via the complex parameter Z'', which depends on the relaxation time. It is represented as the product of an oscillatory function and a convolution integral, which can be interpreted as the evolution of a complex envelope function. The integration of the first term in the bracket of the convolution integral describes the evolution of the burst envelope in a one-dimensional relaxing medium, whereas the integration of the second term describes the evolution of its diffraction correction. The attenuation of the harmonic wave portion of the burst signal does not depend on attenuation of the envelope function $M(t)$. However, the evolution of the complex envelope function represented by the integral in Eq. (14) is affected by the frequency of the oscillation of the harmonic wave portion of the signal due to the modified kernel for the relaxing medium with modified parameters related to the net increase in phase speed and the relaxation time Eq. (10).

It is convenient for analysis purposes to transform the convolution integral in Eq. (14) to a dimensionless form. We introduce the dimensionless retarded time normalized by the relaxation time τ_{relax}, and the dimensionless separation between piston and receiver

$$\tau = t_{\mathrm{ret}}/\tau_{\mathrm{relax}}, \quad \overline{\tau} = s/\tau_{\mathrm{relax}}, \quad \zeta = m'z/2c_\infty\tau_{\mathrm{relax}}, \quad (16)$$

so that

$$Z''(\theta) = \left(\zeta^2 + 4\left(\frac{2c_\infty\tau_{\mathrm{relax}}a}{m'} \right)^2 \cos^2\theta \right)^{1/2}. \quad (17)$$

Then the average pressure Eq. (14) can be written as

$$\langle p(z,t)\rangle = e^{i\omega_0 t - ik(\omega_0)z} \int_{-\infty}^{+\infty} \left[G(\zeta, \tau - \overline{\tau}, \omega_0, \tau_{\mathrm{relax}}) - \frac{4e^{ik(\omega_0)z}}{\pi} \int_0^{\pi/2} G(Z(\theta), \tau - \overline{\tau}, \omega_0, \tau_{\mathrm{relax}}) \times e^{-ik(\omega_0)Z(\theta)} \sin^2\theta\, d\theta \right] M(\overline{\tau})\, d\overline{\tau}, \quad (18)$$

where the dimensionless Green's function $G(\zeta, \tau, \omega_0, \tau_{\mathrm{relax}})$ takes the form

$$G(\zeta,\tau,\omega_0,\tau_{\text{relax}})=\exp\left(-\frac{\zeta}{1+i\tau_{\text{relax}}\omega_0}\right)\left[\delta(\tau)+H(\tau)\times\left(\frac{\zeta}{\tau}\right)^{1/2}\exp[-\tau(1+i\tau_{\text{relax}}\omega_0)]\times I_1[2(\zeta\tau)^{1/2}]\right]. \quad (19)$$

In Eq. (18), the first term in the large brackets is the plane wave portion, while the second term in the brackets reflects the three-dimensional character of the wave.

The diffraction correction integral term in Eq. (18) for the average pressure depends on the slowly varying envelope function $M(\tau)$ and the plane harmonic wave oscillation ω_0. However, it is possible to asymptotically simplify the diffraction correction in the case of the short wave limit by applying Laplace's method for contour integrals (see Ref. 18, Theorem 6.1, p. 125) to the internal integral over a nonzero separation z, comparable with the radius of the piston. This yields an asymptotic series in powers of $\Omega_0=a\omega_0/c_\infty$, calculated in two stationary points $\theta=0$ and $\theta=\pi/2$. Only the point $\theta=\pi/2$ contributes to the leading term in this asymptotic representation. This is a result of the edge of the circular piston dominating the diffraction correction term. At high frequency, $\Omega_0\gg 0$, the asymptotic representation of the average pressure is

$$\langle p(z,t)\rangle\approx\left(1-e^{-i\pi/4}\sqrt{\frac{2z}{\pi k(\omega_0)a^2}}\right)e^{i\omega_0 t-ik(\omega_0)z}\times\int_{-\infty}^{+\infty}G(\zeta,\tau-\overline{\tau},\omega_0,\tau_{\text{relax}})M(\overline{\tau})d\overline{\tau}. \quad (20)$$

Here, the term in parentheses is the three-dimensional diffraction correction, the factor $\exp(i\omega_0 t-ik(\omega_0)z)$ describes the plane wave evolution of harmonic oscillation including the effect of attenuation due to relaxational losses, and the integral describes the evolution of the envelope $M(t)$ of the burst Eq. (2). The form resembles the appearance of the evolution of a single pulse if the frequency of oscillation $\omega_0=0$ and the product $\tau_{\text{relax}}\omega_0$ in the kernel Eq. (19) is set to zero.[11]

If, in addition, the period of harmonic oscillation $1/\tau_{\text{relax}}\omega_0$ is much less than the envelope duration, Eq. (20) can be simplified to

$$\langle p(z,t)\rangle\approx\left(1-e^{-i\pi/4}\sqrt{\frac{2z}{\pi k(\omega_0)a^2}}\right)e^{i\omega_0 t-\alpha z}M(t). \quad (21)$$

Here $\alpha=m'/2c_\infty\tau_{\text{relax}}$, which does not depend on the frequency of modulation ω_0, is the relaxational attenuation at a frequency well above the frequency associated with the maximum relaxational attenuation. Instead of the envelope $M(t)$ being part of a complex integral as in Eq. (20), the envelope in Eq. (21) appears as the initial envelope. The relaxational attenuation of the modulated pulse depends only on $\exp(i\omega_0 t-az)$ and the diffraction correction in parentheses in Eq. (21). Thus, the result is that the envelope $M(t)$ is attenuated by the factors preceding it, but is not distorted in any way.

Another simplification of Eq. (18) is to consider a plane burst rather than a three-dimensional burst in a relaxing medium. In this case, the second term in the brackets in Eq. (18) drops out. Thus,

$$\langle p(z,t)\rangle=e^{i\omega_0 t-ik(\omega_0)z}\times\int_{-\infty}^{+\infty}G(\zeta,\tau-\overline{\tau},\omega_0,\tau_{\text{relax}})M(\overline{\tau})d\overline{\tau}=e^{i\omega_0 t-ik(\omega_0)z}M_{\text{relax}}(\tau). \quad (22)$$

This is an exact equation, which is a generalization of the solution for a single pulse in a relaxing medium[11–16] that allows for a plane burst rather than a single pulse.

III. ANALYSIS AND NUMERICAL EXAMPLES

We analyze evolution of a planar burst in a relaxing medium using the exact expression Eq. (22), in which relaxation and the burst envelope are factors. The convolution integral in Eq. (22) with kernel Eq. (19) describes evolution of a complex burst envelope, $M_{\text{relax}}(\tau)$, of an oscillating pulse in the same form as the evolution of a single pulse.[16] An important distinction from the case of single pulse propagation is the oscillatory nature of the complex kernel as the function of retarded time, depending on the relaxation time and the frequency of oscillation. The interplay of these independent parameters results in an anomalous behavior of the complex envelope, which is quite different from that for a single pulse.

To demonstrate the interplay between the burst frequency and the burst envelope, we calculate the absolute value of the burst envelope for different separations between the piston and the receiver for a variety of oscillation frequencies using a Gaussian envelope $M(t)=\exp(-t^2/2\gamma^2\tau_{\text{relax}}^2)$. The following parameters were used for the calculation: radius of piston $a=0.01$ m, the frozen speed of sound $c_\infty=300$ m/s, relaxation time $\tau_{\text{relax}}=10^{-7}$ s, and the net increase in phase speed $m'=10^{-3}$. These parameters are typical of ultrasonic transducers in air. The parameter γ defines the width of the envelope. The dimensionless separation ζ between the piston and the receiver is defined by Eq. (16).

It is instructive to investigate the evolution of the envelope using Eq. (22) and varying the duration of the Gaussian envelope and dimensionless frequency $\omega_0\tau_{\text{relax}}$. First let the parameter $\gamma=1$ in the Gaussian envelope and let the dimensionless frequency of the oscillations in the burst be $0\leqslant\omega_0\tau_{\text{relax}}\leqslant 20$. In Fig. 1(a), the shape of the transmitted burst and its envelope are shown for $\omega_0\tau_{\text{relax}}=0,0.5,1,2,3,3.5,10,20$, corresponding to 0 to 20 oscillations within the burst envelope. Figure 1(b) shows the evolution of the burst envelope, $M_{\text{relax}}(\tau)$, which corresponds to the integral in Eq. (22). Here the horizontal axis is the dimensionless retarded time $\tau=t_{\text{ret}}/\tau_{\text{relax}}$, and envelope profiles are the calculated at dimensionless distances from the piston of $\zeta=0,2,4,6,8$. The curve for $\zeta=0$ is the Gaussian

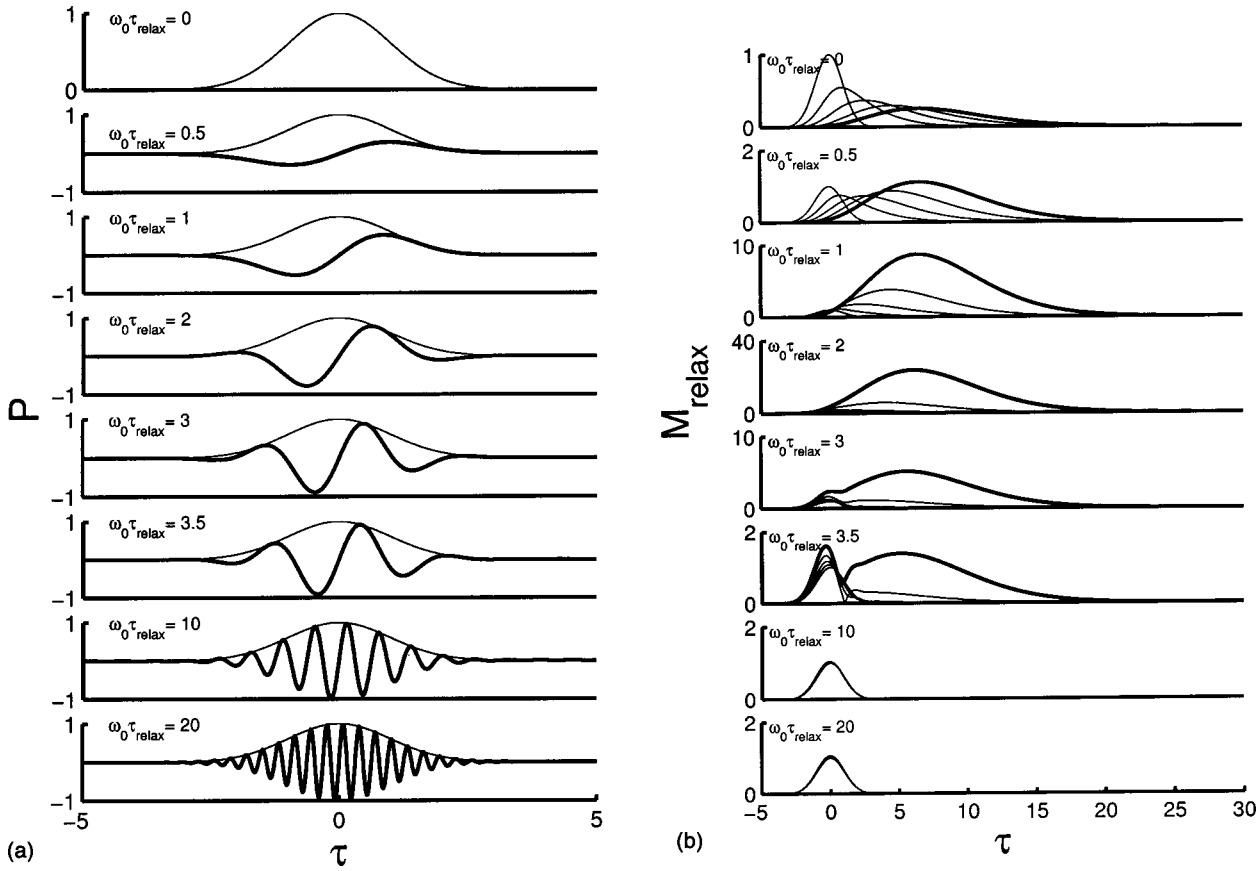


FIG. 1. (a) Gaussian bursts for $\gamma=1$ for different values of the dimensionless parameter $\omega_0\tau_{\text{relax}}=0,0.5,1,2,3,3.5,10,20$. (b) Evolution of the burst envelopes for $\omega_0\tau_{\text{relax}}=0,0.5,1,2,3,3.5,10,20$ at dimensionless separations of $\zeta=0,2,4,6,8$. The Gaussian curve at $\tau=0$ corresponds to $\zeta=0$; the bold curve corresponds to $\zeta=8$ for all cases except $\omega_0\tau_{\text{relax}}=10$, 20 where all curves overlap.

curve centered at $\tau=0$; the curve for $\zeta=8$ is bold; curves for $\zeta=2,4,6$ lie between the $\zeta=0$ and $\zeta=8$ curves.

For $\omega_0\tau_{\text{relax}}=0$, which is the case of a single pulse, the evolution of the complex envelope degenerates to the evolution of the Gaussian single pulse. The pulse amplitude decreases with distance, while the pulse duration increases. Andreev *et al.*[16] observed a similar broadening of a single pulse and a decrease in the pulse amplitude.

For $\omega_0\tau_{\text{relax}}=0.5$, similar envelope broadening occurs, accompanied by decreasing amplitude for short distances from the piston. However, for $\zeta\geqslant 6$ the maximum of the burst envelope increases with distance. Such amplification causes the attenuation of the oscillating pulse to differ substantially from the attenuation of a harmonic wave or a single pulse. The combined effects of the burst oscillation, the burst envelope, and the relaxation of the media produce this result. This amplification can be quite large, noting the different vertical scales used in Fig. 1(b). The amplification reaches a maximum for the case of $\omega_0\tau_{\text{relax}}=2$, where the maximum amplitude of the burst envelope is over a magnitude larger than that of the transmitted burst. For $3\leqslant\omega_0\tau_{\text{relax}}\leqslant 3.5$, the amplification weakens and the envelope splits into in two humps. The weakening may be attributed to the oscillation of the second term in the kernel Eq. (19) as a function of τ, which leads to degradation of its convolution with the burst envelope and eventual domination of the convolution with the first term. As the oscillation frequency increases further, the amplification decreases, so that for $\omega_0\tau_{\text{relax}}=10$, 20 the amplification is negligible and the burst duration is unchanged because the oscillating pulse attenuates as a plane harmonic wave.

The maxima of the envelope magnitudes are shown in Fig. 2 as functions of the dimensionless parameter $\omega_0\tau_{\text{relax}}$ for distances from the piston of $\zeta=0,2,4,6,8$. The horizontal

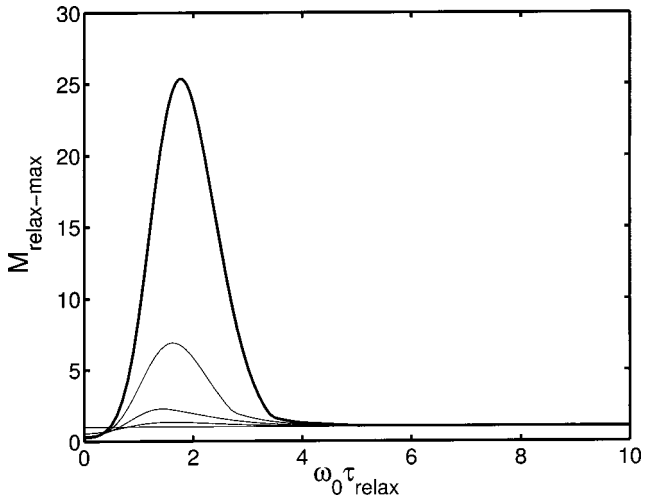


FIG. 2. Maximum values of the Gaussian burst envelopes calculated at dimensional separations of $\zeta=0,2,4,6,8$ for $\gamma=1$. The horizontal line at $M_{\text{relax\_max}}=1$ corresponds to $\zeta=0$; the bold curve corresponds to $\zeta=8$.

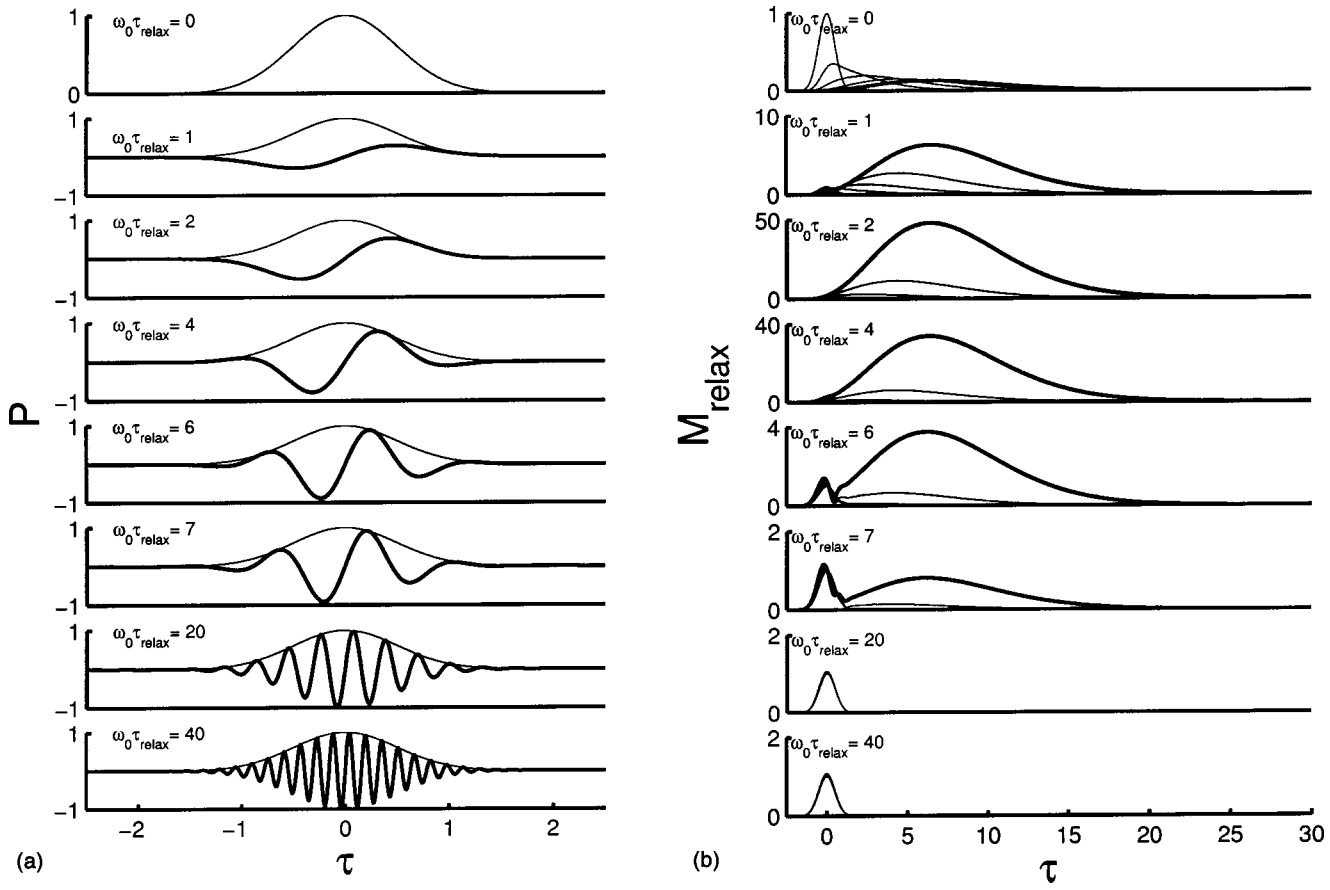


FIG. 3. (a) Gaussian bursts for $\gamma=0.5$ for different values of the dimensionless parameter $\omega_0\tau_{\text{relax}}=0,1,2,4,6,7,20,40$. (b) Evolution of the burst envelopes for $\omega_0\tau_{\text{relax}}=0,1,2,4,6,7,20,40$ at dimensionless separations of $\zeta=0,2,4,6,8$. The Gaussian curve at $\tau=0$ corresponds to $\zeta=0$; the bold curve corresponds to $\zeta=8$ for all cases except $\omega_0\tau_{\text{relax}}=20$, 40 where all curves overlap.

line at $M_{\text{relax\_max}}=1$ corresponds to $\zeta=0$; the bold curve corresponds to $\zeta=8$. The magnitude of the envelope grows quite large as the separation distance ζ increases for $1\leqslant\omega_0\tau_{\text{relax}}\leqslant 3$.

Now we compare this result for $\gamma=1$ with the evolution of the Gaussian envelope for shorter and longer durations, $\gamma=0.5$ and $\gamma=2$. In Fig. 3(a) for $\gamma=0.5$, half as many oscillations are included in the shorter burst envelope than for the duration $\gamma=1$ at the same $\omega_0\tau_{\text{relax}}$. In Fig. 3(b), the amplification of the envelope occurs over a wider range of the dimensionless frequency, $1\leqslant\omega_0\tau_{\text{relax}}\leqslant 6$. In addition, the maximum amplitudes are higher than those for a wider envelope, $\gamma=1$ in Fig. 1(b). This is also reflected in Fig. 4, when compared to Fig. 2. The maximum amplification can be more than twice as much for the shorter envelope, noting the differences in the vertical scales in the figures.

For $\gamma=2$, more oscillations are included within the longer envelope than for $\gamma=1$ at the same $\omega_0\tau_{\text{relax}}$ as shown in Fig. 5(a), noting the different horizontal scale from previous similar figures. Like the previous two cases, amplification of the envelope occurs as shown in Fig. 5(b), but in this case the range of frequencies is smaller, $0.5\leqslant\omega_0\tau_{\text{relax}}\leqslant 2$. In addition, the maximum amplitudes are much smaller. These results are more evident comparing Fig. 6 to Figs. 2 and 4, noting the substantially different vertical scales.

In spite of the differences in the three cases, the amplification always occurs when $\omega_0\tau_{\text{relax}}$ has an order of magnitude of unity. However, the degree of amplification and the frequency at which the maximum occurs both increase as the envelope shortens. It appears that when the period of the pulse is similar to the duration of the envelope, the amplification occurs. For example, the combinations ($\omega_0\tau_{\text{relax}}=4$, $\gamma=0.5$), ($\omega_0\tau_{\text{relax}}=2$, $\gamma=1$), and ($\omega_0\tau_{\text{relax}}=1$, $\gamma=2$) all consist of one full wave within the burst envelope [see Figs.

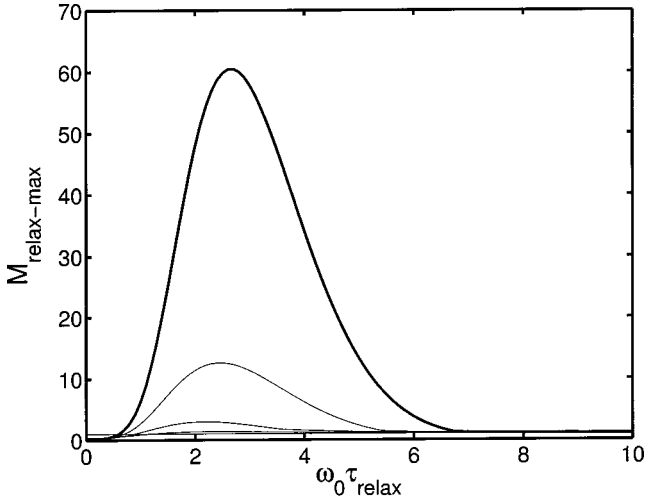


FIG. 4. Maximum values of the Gaussian burst envelopes calculated at dimensional separations of $\zeta=0,2,4,6,8$ for $\gamma=0.5$. The horizontal line at $M_{\text{relax\_max}}=1$ corresponds to $\zeta=0$; the bold curve corresponds to $\zeta=8$.

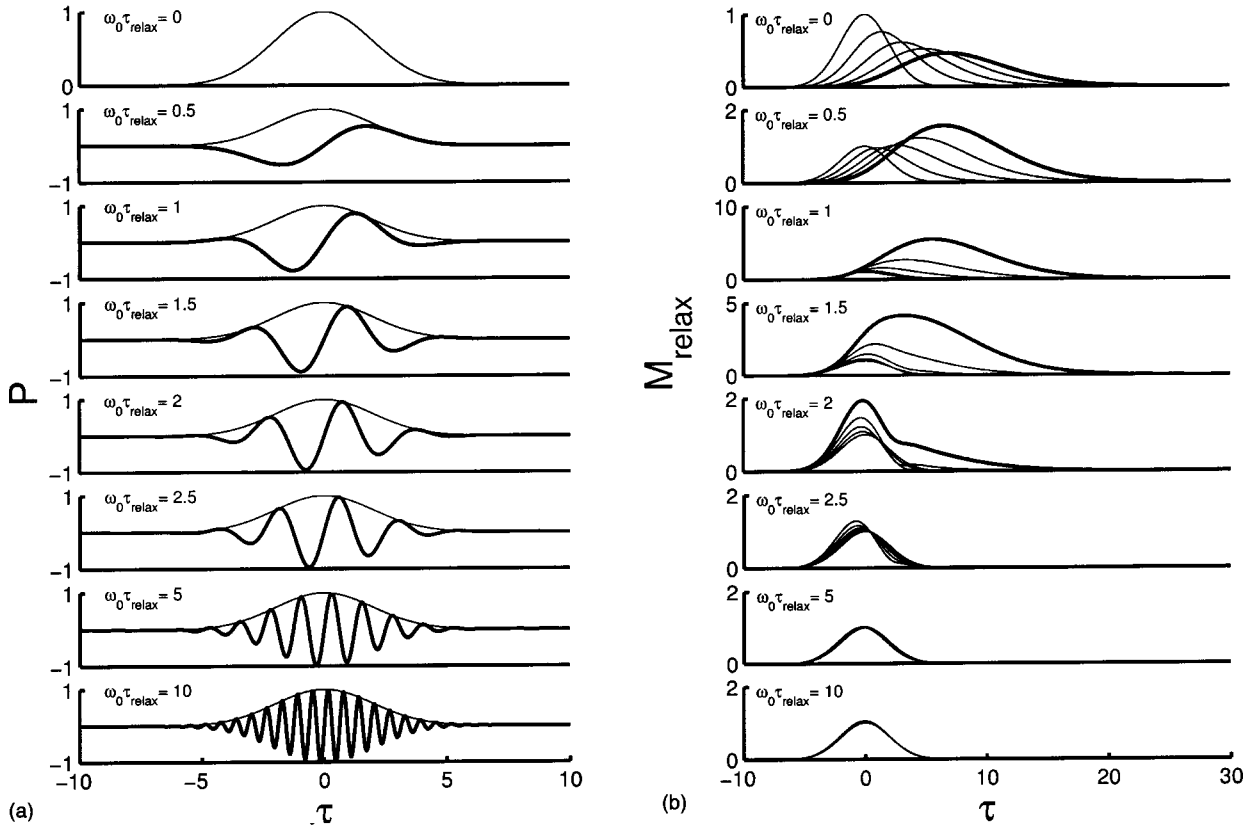


FIG. 5. (a) Gaussian bursts for $\gamma=2$ for different values of the dimensionless parameter $\omega_0\tau_{relax}=0,0.5,1,1.5,2,2.5,5,10$. (b) Evolution of the burst envelopes for $\omega_0\tau_{relax}=0,0.5,1,1.5,2,2.5,5,10$ at dimensionless separations of $\zeta=0,2,4,6,8$. The Gaussian curve at $\tau=0$ corresponds to $\zeta=0$; the bold curve corresponds to $\zeta=8$ for all cases except $\omega_0\tau_{relax}5,10$ where all curves overlap.

1(a), 3(a), and 5(a)]. In all three cases, the maximum amplification occurs very near these frequencies, as shown in Figs. 2, 4, and 6.

The nature of the amplification of the burst envelope can be understood by looking into dependence of the kernel Eq. (19) on the dimensionless frequency $\omega_0\tau_{relax}$. The kernel, G, is a product of two factors: the first one, $\exp(-\zeta/(1+i\omega_0\tau_{relax}))$, depends on the distance from the piston; the second factor, which is in brackets in Eq. (19), is the convolution with the Gaussian envelope. Its second term oscillates as $\exp(-i\omega_0\tau_{relax})$. When the period of Gaussian envelope is comparable with the period of oscillation, the second term in the brackets dominates, and the convolution of the oscillating term with $M(\tau)$ is an increasing function of distance ζ. Consequently, the maximum amplitude of the envelope increases with distance. However, the rate of the growth of the envelope's amplitude is modulated by the factor in the kernal preceding the bracketed term, $\exp(-\zeta/(1+i\omega_0\tau_{relax}))$. Of course, for larger $\omega_0\tau_{relax}$, this term is less significant. Thus, for small γ, where amplification occurs for a larger value of $\omega_0\tau_{relax}$, the envelope is amplified more than for large γ, where the amplification occurs at a smaller value of $\omega_0\tau_{relax}$.

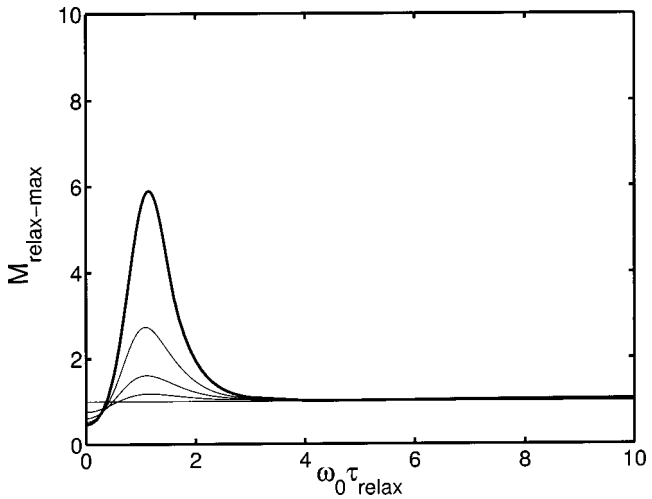


FIG. 6. Maximum values of the Gaussian burst envelopes calculated at dimensional separations of $\zeta=0,2,4,6,8$ for $\gamma=2$. The horizontal line at $M_{relax\_max}=1$ corresponds to $\zeta=0$; the bold curve corresponds to $\zeta=8$.

From Figs. 2, 4, and 6 there appears to be frequency at which there is a change from amplification to a situation of nearly no amplification, which is characteristic of a harmonic wave. This occurs when $\omega_0\tau_{relax}$ is sufficiently high so that the oscillation of the second term in the kernel yields a convolution integral with very small magnitude, and the exponential factor $\exp(-\zeta/(1+i\omega_0\tau_{relax}))$ dominates the evolution of the complex envelope.

IV. CONCLUDING REMARKS

Equation (18) describes the interaction between the diffractional attenuation and the relaxational attenuation as a tone burst propagates between circular transducers in a relaxing medium. Applying this formulation to the more simple case of a planar burst demonstrates the interaction

between the burst duration, the harmonic frequency, and the relaxation time constant. When the duration of the burst coincides with the relaxation time and the period of the wave, amplification of the burst envelope occurs. When the duration of the burst is substantially longer than the period of the wave the attenuation is similar to that for a harmonic wave. Provided that the relaxation time is known, Eq. (18) provides a means for determining the combined effect of diffraction and relaxation on the attenuation of a tone burst.

ACKNOWLEDGMENTS

This work was supported by DOE under subcontract to Commercial Electronics, Inc., Broken Arrow, OK, and by the Ford Motor Company. We thank Dr. Jaco Visser and Scott Phillips for several helpful discussions.

[1] E. P. Papadakis, ''Ultrasonic diffraction from single apertures with application to pulse measurements and crystal physics,'' in *Physical Acoustics: Principles and Methods*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1975), Vol. XI, pp. 151–211.

[2] H. Seki, A. Granato, and R. Truell, ''Diffraction effects in the ultrasonic field of a piston source and their importance in the accurate measurements of attenuation,'' J. Acoust. Soc. Am. **28**, 230–238 (1956).

[3] E. Lommel, ''Die Beugunserscheinungen einer kreisrunden Oeffnung,'' (in German) Abh. Bayer. Akad. Wissensch. **15**, 233 (1886). Lommel's work is discussed in English in A. Gray, G. B. Mathews, and T. M. MacRobert, *A Treatise on Bessel Functions and Their Application to Physics* (Dover, New York, 1966), Chap. 14.

[4] A. O. Williams, ''The piston source at high frequencies,'' J. Acoust. Soc. Am. **23**, 1–6 (1951).

[5] A. S. Khimunin, ''Numerical calculation of the diffraction corrections for the precise measurement of ultrasound absorption,'' Acustica **27**, 173–181 (1972).

[6] A. S. Khimunin, ''Numerical calculation of diffraction corrections for precise measurement of ultrasound phase velocity,'' Acustica **32**, 192–200 (1975).

[7] P. H. Rogers and A. L. VanBuren, ''An exact expression for the Lommel diffraction correction integral,'' J. Acoust. Soc. Am. **55**, 724–728 (1974).

[8] H. O. Kneser, ''Theory of dispersive waves,'' Ann. Phys. (Leipzig) **11**, 771–777 (1931) (in German).

[9] L. I. Mandel'shtam and M. A. Leontovich, ''On the theory of sound absorption,'' Zh. Eksp. Teor. Fiz. **7**, 439–444 (1937) (in Russian).

[10] Y. Dain and R. M. Lueptow, ''Acoustic attenuation in three-component gas mixtures—theory,'' J. Acoust. Soc. Am. **109**, 1955–1974 (2001).

[11] L. A. Vainshtein, ''Propagation of pulses,'' Sov. Phys. Usp. **19**, 189–205 (1976).

[12] S. Z. Dunin, ''Wave propagation in a weakly dispersive medium,'' J. Appl. Mech. Tech. Phys. **27**, 124–127 (1997).

[13] S. Z. Dunin and G. A. Maksimov, ''Propagation of pulses in media having a spectrum of relaxation times,'' Sov. Phys. Acoust. **34**, 601–604 (1988).

[14] V. G. Andreev, O. A. Sapozhnikov, and S. T. Timofeev, ''Evolution of acoustic pulses in a relaxational medium,'' Acoust. Phys. **40**, 172–175 (1994).

[15] V. A. Larichev and G. A. Maksimov, ''Analytical description short pulse propagating dynamics in a relaxation medium,'' Acoust. Phys. **43**, 314–321 (1997).

[16] V. G. Andreev, Y. A. Pishchal'nikov, O. A. Sapozhnikov, V. A. Khokhlova, and R. O. Cleveland, ''Diagnostics of a relaxing medium by an acoustic pulse with shock front,'' Acoust. Phys. **45**, 9–13 (1999).

[17] O. V. Rudenko and S. I. Soluyan, *Theoretical Foundations of Nonlinear Acoustics* (Plenum, New York, 1977), pp. 79–96.

[18] F. W. J. Olver, *Asymptotics and Special Functions* (Peters, Wellesley, MA, 1997), p. 125.

Inverse ultrasonic determination of imperfect interfaces and bulk properties of a layer between two solids

A. Baltazar, L. Wang, B. Xie, and S. I. Rokhlin[a)]
Nondestructive Evaluation Program, Edison Joining Technology Center, The Ohio State University, 1248 Arthur E. Adams Drive, Columbus, Ohio 43221

(Received 5 March 2003; revised 8 June 2003; accepted 23 June 2003)

A method for determination of the complete set of physical, geometrical, and interfacial properties of an isotropic layer embedded between two known solids is discussed. These properties are: Lamé elastic moduli, density and thickness of the layer, and complex normal and transverse interfacial stiffnesses between the layer and the substrates. The properties are combined in the form of eight nondimensional parameters, which are determined from experimental reflection spectra at two incident angles: normal and oblique. The conditions for simultaneous determination of bulk layer properties and the interface normal and transverse springs with losses and the stability of the inversion method against data scatter are addressed. The inversion model is validated by experiment on normal and angular ultrasonic reflectivity from a layer between two semispaces in dry mechanical contact and from an environmentally degraded adhesive joint. The layer properties were measured independently, showing good agreement with the reconstructed results. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1600723]

PACS numbers: 43.35.Cg, 43.20.El, 43.35.Zc [YHB]

I. INTRODUCTION

Modern materials are often engineered as layered structures such as thin films in electronic parts, coatings for service life extension, and advanced composite laminates. Also, different parts are often joined using thin layers, for example, adhesive bonding, brazing, and diffusion bonding. Properties of thin layers are often tailored to satisfy design requirements and thus should be measured as part of quality process control. The bond integrity and strength (adhesion) between the layer and the substrate is most difficult to control during manufacturing and is difficult to measure nondestructively. However, it is well known that the strength and life expectancy of many engineering structures depends critically on the bond integrity between structural components. Imperfections along the bondline such as cracks, porosity, inclusions, etc. can significantly degrade the performance of the joint. Very often the imperfections are confined to a very thin layer in the form of an interphase separating the joining materials which is difficult to characterize. Ultrasonic methods have been very promising for nondestructive evaluation of such interphases.

When the interphasial layer between the solids is thin and the imperfections are flat, the interphasial layer can be modeled as an infinitely thin interface connecting the solids by distributed springs to account for interfacial stiffness reduction due to imperfections; and analytically, such an interface can be modeled using spring boundary conditions[1] where spring values have been related by a model to the microcrack distribution.[1,2] Significant effort has been put into experimental and theoretical studies of ultrasonic wave interaction with imperfect interfaces.[1–24] The goal of these studies was to characterize interface imperfections measuring ultrasonic signatures perturbed by the interface. Both bulk and guided waves were used for this purpose. The reflection response from an elastically perfect interface is frequency independent while from an interface with spring boundary conditions it depends on frequency. By measuring this frequency response one can determine the interfacial spring stiffness. For example, in more recent studies Rokhlin *et al.*[18] used the reflectivity method to characterize an imperfect interface between a plasma spray coating and a substrate, and Guo *et al.*[19] applied high-frequency line-focus acoustic microscopy to characterize an imperfect interface between a thin film and a substrate. In these papers models have been developed and applied in the inversion procedure to determine interfacial spring constants from the measured acoustic signatures.

Most previous work has focused on single imperfect interface characterization. However, characterization of imperfect interfaces between a layer and two substrates is also important for many material joining and coatings applications, for example for adhesive bonds which deteriorate in a humid environment predominantly along the adhesive/adherend interface[20] where a model with water-filled microdisbonds was developed to describe damage by distributed transverse springs and to characterize the degradation. The distinctive feature of this type of material system is that the layer is thin, and the ultrasonic signals reflected (normally or obliquely) from the front and back sides of the layer are not separated in the time domain and interfere. Thus, ultrasonic spectroscopy may be applied to characterize two imperfect interfaces separated by a layer of thickness h. This problem was addressed recently by Lavrentyev and Rokhlin,[21] who used ultrasonic spectroscopy to evaluate the interfacial conditions from the spectra of longitudinal and shear waves reflected normally by the neighboring interfaces. They showed that the frequencies at which the reflected spectrum minima

a)Electronic mail: rokhlin.2@osu.edu

occur are sensitive to interfacial conditions, and, therefore, may be used for their monitoring.

More recently, Baltazar *et al.*[22,23] demonstrated experimentally that the values of the interfacial normal and transverse stiffness constants obtained from data acquired at normal incidence may be used to predict the dynamic response of the interface even when the interrogating wave impinges on the interface at oblique incidence, i.e., these two interfacial stiffnesses describe the imperfect interface uniquely. They also provided a theoretical description of the micromechanics of rough, contacting surfaces under static normal and shear loads. In particular, they predicted the values for the ratio between the normal and transverse interfacial stiffness constants.

In our and other previous studies it was assumed that the properties of the bondline layer are known and only imperfect interface properties need to be determined from the experiment. However, in many cases the properties of the bonding layer are also unknown. If the ultrasonic reflectivity signature from such a layer has an anomaly it is important to know if it is due to the interface (adhesion) imperfections or due to the layer properties anomaly. Thus, an inverse-independent determination of these properties from the experimental ultrasonic signature is important. In this paper we address this problem both theoretically and experimentally. The development is an extension of the inversion method proposed by Lavrentyev and Rokhlin[24] and extended by Wang *et al.*,[25] who performed reconstruction of all geometrical and acoustic properties of a layer perfectly embedded between substrates.

II. INVERSION MODEL FOR THE RECONSTRUCTION OF INTERFACIAL AND BULK LAYER PROPERTIES

A. Definition of independent parameters for embedded layer with imperfect boundary conditions

Let us consider a solid layer imperfectly embedded between two solids as shown in Fig. 1(a). To characterize the interfaces and layer properties the normal and oblique ultrasonic reflection signatures are taken from the layer. Reconstruction of the interfaces and the layer parameters is model based. In the model the imperfect interface conditions between the layer and the substrates are represented by distributed complex normal and shear springs as shown in Fig. 1(a). The spring model is quasistatic and valid for wavelengths larger than the dimensions of the interface discontinuities. Using this concept the boundary conditions between the layer and the substrates are described as

$$\begin{aligned}\sigma_{zz}(z=\pm h/2+0^{\pm}) &= \sigma_{zz}(z=\pm h/2+0^{\mp}) \\ &= K_N[u_z(z=\pm h/2+0^{+}) \\ &\quad - u_z(z=\pm h/2+0^{-})], \\ \sigma_{xz}(z=\pm h/2+0^{\pm}) &= \sigma_{xz}(z=\pm h/2+0^{\mp}) \\ &= K_T[u_x(z=\pm h/2+0^{+}) \\ &\quad - u_x(z=\pm h/2+0^{-})].\end{aligned} \tag{1}$$

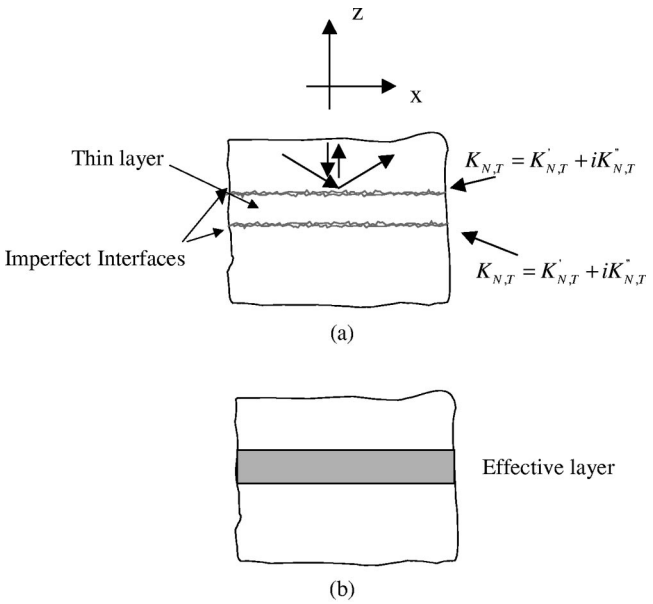


FIG. 1. (a) Layer embedded between two solids with an imperfect interface consisting of two rough surfaces. Using spring approximation complex normal K_N (N/m^3) and transverse K_T springs replace imperfect interface. (b) Model with effective layer perfectly bonded to substrates. The layer has same thickness but effective elastic moduli which are selected to match as closely as possible the reflected ultrasonic signatures.

Equation (1) represents continuity of stress σ_{ik} and discontinuity of displacement u_i at the interfaces; h is the layer thickness. The properties of the interfacial contacts are described by normal K_N (N/m^3) and transverse K_T complex interfacial spring constants. At perfect (welded) contacts between solids the spring constants approach infinity.

We consider the layer properties and the interfacial springs, in the boundary conditions (1), to be unknown. We assume that the substrate properties are known *a priori* and the interfacial springs are the same on both interfaces between the layer and the substrates. We use six parameters[24,25] to describe all unknown variables: bulk layer properties and two interfacial spring constants. Those parameters will be reconstructed from measured normal and angular reflection ultrasonic spectra. The model includes in general ten parameters: Lame's constants λ and μ, thickness h, density ρ, longitudinal α_l and shear α_t attenuations, and the interfacial normal and transverse complex spring constants

$$\lambda,\ \mu,\ h,\ \rho,\ \alpha_l,\ \alpha_t,\ K_N = K'_N + iK''_N,\ K_T = K'_T + iK''_T. \tag{2}$$

However, as discussed below, the attenuation in the layer and the imaginary parts in the spring constants cannot be separated, and therefore a total of eight parameters will be considered to be unknown. Generalizing the procedure described in Refs. 24 and 25, from the parameters (2) we defined two groups of nondimensional parameters. From the normal-reflection spectrum the following four nondimensional parameters will be determined:

$$Z_n = \frac{Z_l}{Z_s},\quad H_l = \frac{h\omega_0}{V_l},\quad \overline{K_N} = K_N/(\omega_0 Z_l),\ \alpha_l, \tag{3}$$

where $Z_{l,s}=\rho_{o,s}V_{l,sl}$ is the acoustic impedance; the subscript s denotes quantities belonging to the substrate; ρ_o is the density of the adhesive and ρ_s is substrate density, V_l is the longitudinal velocity in the adhesive, and V_{sl} that in the substrate; $\overline{K_N}$ is the nondimensional complex normal interfacial spring constant, and α_l is the longitudinal attenuation in the thin layer. The angular frequency was chosen as $\omega_0=1\ \mu s^{-1}$ for convenience. At oblique incidence an additional set of four nondimensional parameters is determined

$$H_{t\theta}=\frac{V_l}{V_{sl}},\quad H_{l\theta}=\frac{V_t}{V_{st}},\quad \overline{K_T}=K_T/(\omega_0 Z_t),\quad \alpha_t. \tag{4}$$

Here, $\overline{K_T}$ is the complex nondimensional transverse interfacial spring constant, V_t is the shear velocity in the adhesive and V_{st} that in the substrate; $Z_t=\rho V_t$, and α_t is shear attenuation in the adhesive.

The energy loss at the interface is accounted for by imaginary parts in the interfacial stiffness constants ($K_{N,T}=K'_{N,T}+iK''_{N,T}$). Physically, the interfacial losses at the interfaces depend on the ultrasonic scattering. Our analysis shows that the imaginary parts in the complex springs produce the same effect on the reflection spectra as attenuation in the thin embedded layer α_l, α_t and therefore they cannot be separated by our method. Such separation may be possible if one considers a different frequency dependence of these variables; however, this will require a much wider frequency bandwidth of measurement than is used in our work. Thus, depending on the physical conditions we assume that only one of the pair of attenuations is unknown and used for reconstruction: either the attenuation coefficients α_l, α_t inside the thin layer or the imaginary parts of the interfacial spring constants; the other pair is considered to be known. We further discuss this issue in Sec. IV.

B. Discussion of separation of interfacial from bulk layer properties

As demonstrated in Refs. 24 and 25, all properties of the layer perfectly bonded to the substrates ($K'_N, K'_T=\infty$) can be reconstructed from the normal and oblique reflectivity spectra. For this reason, to describe the imperfection of the interfaces (degradation) the effective layer model [Fig. 1(b)] was used.[26] In this model the imperfectly embedded layer is replaced by an effective (equivalent) which is perfectly bonded to the substrates. The layer has the same thickness, but with effective elastic moduli which are selected to match as closely as possible the reflected ultrasonic signatures; the effective moduli (ultrasonic velocities) will be smaller than the actual ones to account for the reduction of the interface stiffness. Adding interfacial properties as unknown variables introduces additional unknown parameters to be determined. Assuming that normal and angular reflection spectra are measured, the first question to be addressed is: is the separation of interfacial from bulk properties of the layer possible in principle? In other words, is the problem unique or can the identical ultrasonic reflectivity signature be obtained from two different models: the first is the effective layer model with a perfectly bonded layer between the substrates and

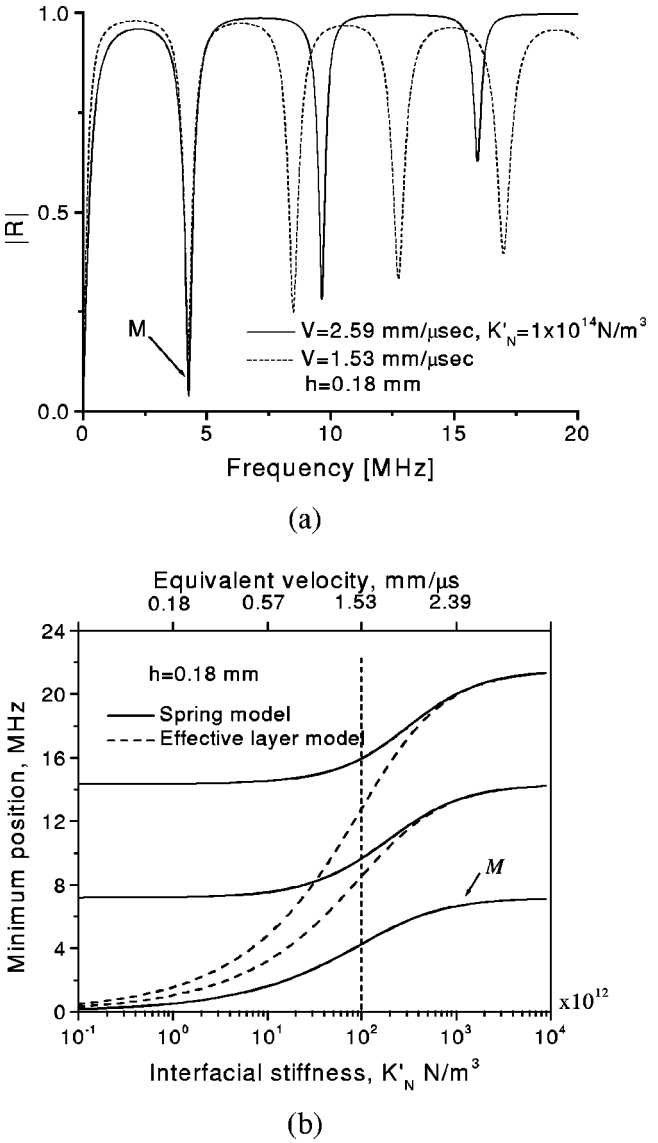


FIG. 2. Effect of bulk and interfacial properties on the spectrum of the reflected signal from the embedded epoxy layer between aluminum substrates; (a) Reflection coefficient |R| versus frequency. The solid line is for a thin epoxy layer embedded between two aluminum plates with imperfect interfaces ($K'_N=1\times10^{14}$ N/m^3). The dashed line is for the effective layer model with perfect interfaces and the reduced bulk longitudinal wave velocity; it has been determined by matching of the reflection spectrum at first minimum (marked M) of the layer with imperfect interfaces (dashed line). (b) Shows by solid lines the shift of the reflection spectra minima versus interfacial spring stiffness ($V_L=2.59$ mm/μs, $h=0.18$ mm). The dashed lines correspond to the effective-properties model as in (a).

effective (artificial) properties, and the second is one with actual layer properties but imperfect bonding to the substrates.

To illustrate this issue, we will use, as an example, simulated reflection spectra at normal incidence. Figure 2(a) shows simulated reflection frequency spectra for layers with two different longitudinal velocities assuming perfect and imperfect interfaces. The solid line is for a thin epoxy layer ($h=0.18$ mm, $V_L=2.59$ mm/μs) embedded between two aluminum plates with imperfect interfaces ($K'_N=1\times10^{14}$ N/m^3). The dashed line is calculated assuming perfect interfaces and the reduced bulk longitudinal wave speed

to $V_L = 1.53$ mm/μs; it has been determined by inverse least-squared optimization of the matching of the reflection spectrum at first minimum (marked M) of the layer with imperfect interfaces (dashed line). We call the model, with such matched properties of a perfectly bonded layer, the effective layer model. As can be seen, only the first minimum position is overlapped for these two cases; other minima cannot be matched. The frequencies of the higher resonance minima become significantly different for the two types of interfacial conditions. Figure 2(b) shows by solid lines the shift of the reflection spectra minima versus interfacial spring stiffness ($V_L = 2.59$ mm/μs, $h = 0.18$ mm). The dashed lines correspond to the effective-properties model. As in Fig. 2(a), to obtain the effective layer velocity the least-squares spectra inversion procedure was used to match the first resonance minimum by the two models at each corresponding value of the interfacial springs [marked M in Fig. 2(b)]. Thus, for both models the curves for the lowest resonance coincide. The positions of the second and third minima are the same for large values of interfacial stiffness (i.e., $K_N' > 10^{15}$ N/m^3); however, they become different as the interfacial stiffness is reduced.

This indicates that in the frequency band covering several layer resonances the reflected spectra change differently with change of bulk layer properties or change of the interfacial properties, and the effective layer model will not match the reflected spectra for imperfect interface conditions in a wide frequency range. On the other hand, performing measurements in a broad frequency range would permit decoupling the effect of the interfacial stiffness from the layer bulk properties. However, if only one spectral minimum is measured the effective layer model and the complete model with the interfacial springs give identical inversion results and the interfacial springs cannot be separated from the layer bulk properties.

The reflection spectra resonances are formed due to multiple reflections of the ultrasonic signal in the layer, and therefore they are sensitive to the interfacial conditions on both interfaces. We do not intend to separate the conditions on these interfaces and assume them to be identical (such a separation is unlikely to be possible by this method). In our experiment the surface preparation of the sample to form the two interfaces between the layer and the substrates is identical; therefore, our assumption that the interfacial springs on both interfaces are identical is reasonable. Our simulations show that when the interfacial properties are not symmetrical the reflection spectra behaves somewhat like that for the symmetrical interfaces with averaged compliances.

C. Inversion algorithm

As described above, the reflection spectra depend on the two sets of nondimensional parameters defined by Eqs. (3) and (4) which are reconstructed from the experimental normal and oblique reflection spectra by minimizing the matching error between the measured and simulated spectra. A multilayered wave propagation algorithm[24,25] with spring boundary conditions between layers is used to compute the reflection spectra in the model. The experimental reflection spectrum from the imperfectly embedded layer used in the reconstruction is deconvolved with the spectra of the reference signal. By this deconvolution, frequency and beam effects of the transducers are removed from the experimental spectra. If the experimental conditions are the same for both the specimen and reference signal measurements, the beam effect due to the finite size of the transducer aperture is removed and the plane-wave approximation may be employed in the properties reconstruction.[25]

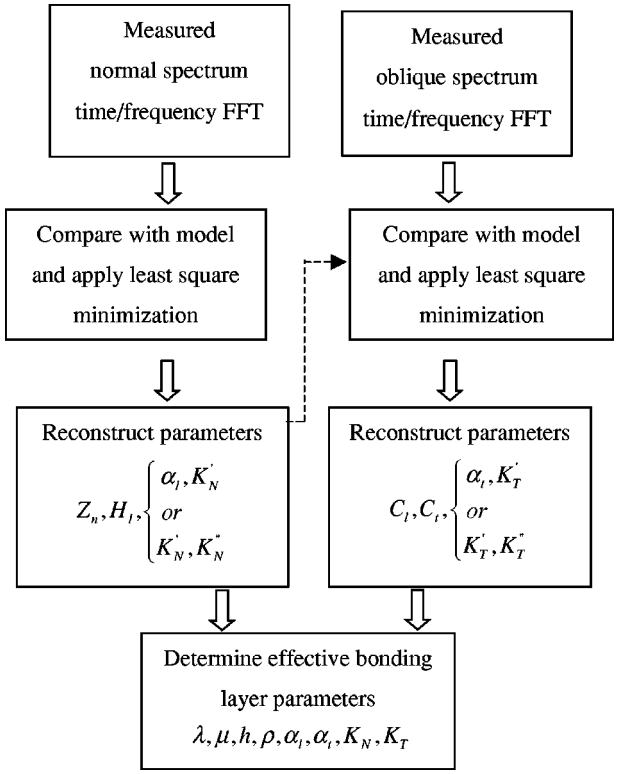


FIG. 3. Reconstruction procedure.

The flow chart for the reconstruction procedure for the parameters (3) and (4) is shown in Fig. 3. First, the normal and oblique spectra are measured experimentally. Then, the parameters defined in Eq. (3) are obtained from the normal spectrum using a least-square optimization algorithm. These reconstructed parameters are then used in a similar optimization procedure to obtain the rest of the parameters in Eq. (4) from the oblique reflection spectrum. After all nondimensional parameters have been determined, the dimensional parameters (2) are obtained from (4) and (5).[22] Considering the unknown nondimensional parameters as a variable vector $\mathbf{X}$ in a four-dimensional space, their values are calculated by least-squares minimization of the sum of squared deviations between the experimental $R_e(\mathbf{X}_0)$ and simulated $R_t(\mathbf{X})$ spectra

$$\mathrm{Error}(\mathbf{X}) = \min \sum_{f=f_1}^{f_2} (|R_t(\mathbf{X})| - |R_e(\mathbf{X}_0)|)^2. \tag{5}$$

$\mathbf{X}$ is an iterated set of four unknown nondimensional parameters [Eq. (3) or (4)], $\mathbf{X}_0$ is the corresponding actual set of material properties (interfacial and bulk layer properties). f_1 and f_2 are the bounds of the frequency band in which mini

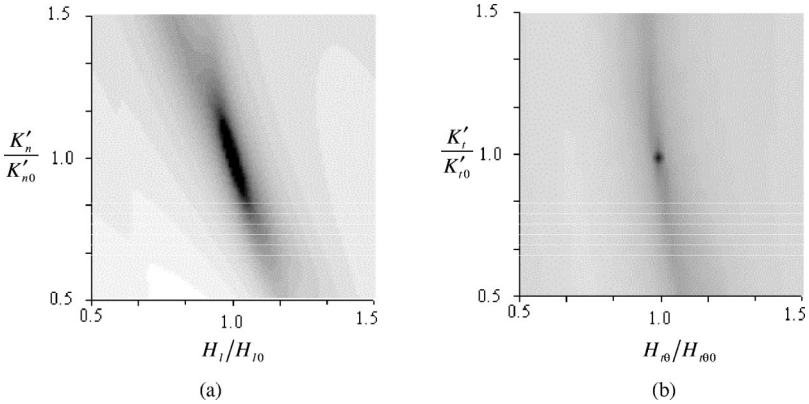


FIG. 4. Cross sections of the 4D error function: (a) $K'_n - H_l$ cross section for normal incidence measurements. (b) $K'_t - H_{t\theta}$ cross section for oblique incidence measurements.

mization is performed. The frequency bounds are selected based on the transducer bandwidth and the resonance frequencies of the adhesive layer. Equation (5) is applied separately to normal and oblique cases as specified in Fig. 3.

The least-square algorithm may converge to one of the minima of the error function (5) in the four-dimensional parameter space. The initial guesses and accuracy of the measurement will affect the convergence of the algorithm and may make it reach a local minimum rather than the global minimum corresponding to the actual parameters. To investigate the influence of the initial guesses we calculated the error function surface. As an example, two cross sections of the error function in the form of gray level images are plotted in Fig. 4. Black corresponds to the smallest level. The nondimensional parameters (K_n and H_l) in Fig. 4(a) and (K_t and $H_{t\theta}$) in Fig. 4(b) are normalized and allowed to vary from 50% to 150% of the actual parameters K_{n0}, H_{l0}, K_{t0}, $H_{t\theta 0}$. Other parameters are kept at the actual value. As can be seen, only one main minimum exists within this range of variation. This indicates that the least square minimization process (5) converges to the actual parameters if the initial guesses are selected within the above variation range and there is no data scatter. Contours of error magnitudes form ellipses, indicating that the rate of convergence depends on the direction of search and the initial guesses. Similar results were obtained for all other parameters.

The inverse problem discussed above depends on the measurement accuracy of both the normal and angular reflection spectra. To assess the effect of data scatter we employed the algorithm described by Wang *et al.*,[25] who used a ''noisy spectrum'' obtained by introducing random scatter into the time-domain signal calculated by the beam model which emulates experimental conditions in the angle beam measurements.[25] The simulated experimental signal has been mixed with random ''noise'' data. The amplitude of random data scatter is 2% of the signal peak amplitudes. Due to data scatter the reconstructed set of layer properties is not exactly equal to the original set. The difference serves as a measure of the effects of the data scatter level and the quality of initial guesses on the inversion algorithm. As an example, in Figs. 5(a) and (b), we plot the relative error between the reconstructed parameter values and the exact original values as histograms for 200 times repeated reconstructions for the same data scatter (2%) and different initial guesses in the range of 20% off the actual values. Figure 5(a) shows the scatter effect on H_l and Fig. 5(b) for $H_{l\theta}$. The standard deviation for H_l is within 0.5% and for $H_{l\theta}$ is 1%.

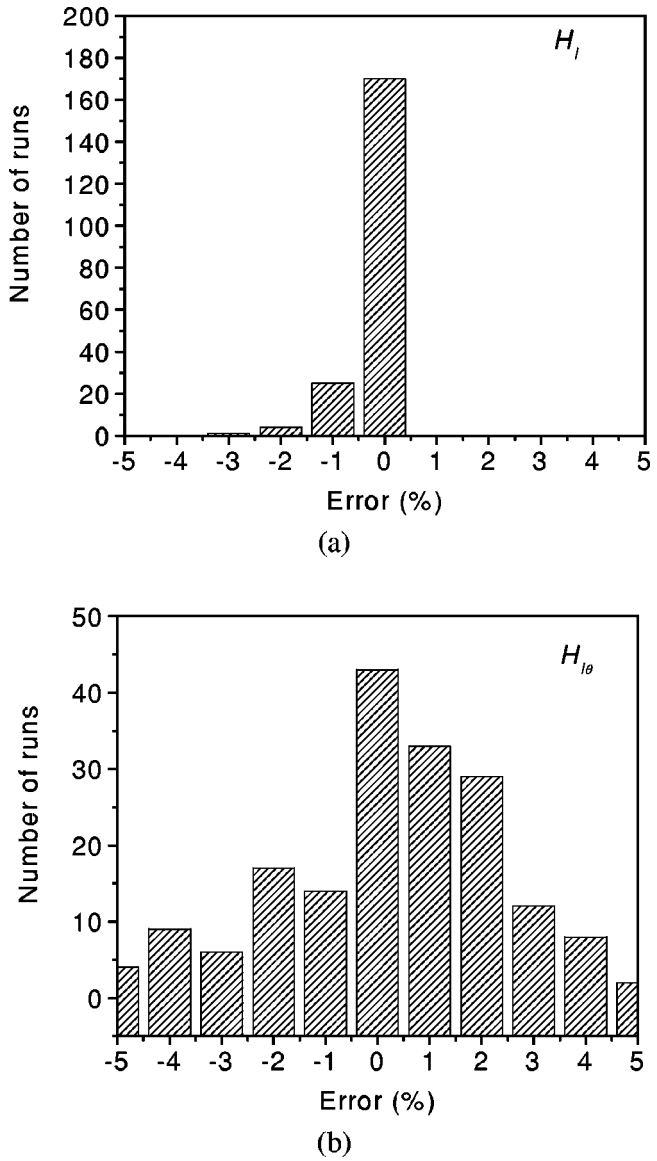


FIG. 5. Histograms for repeated 200 times reconstructions with synthetic ''noisy'' time domain signals (2% scatter). Initial guesses are 20% off the actual values. (a) Results for H_l for normal incidence measurement; (b) result for $H_{l\theta}$ for oblique incidence measurement.

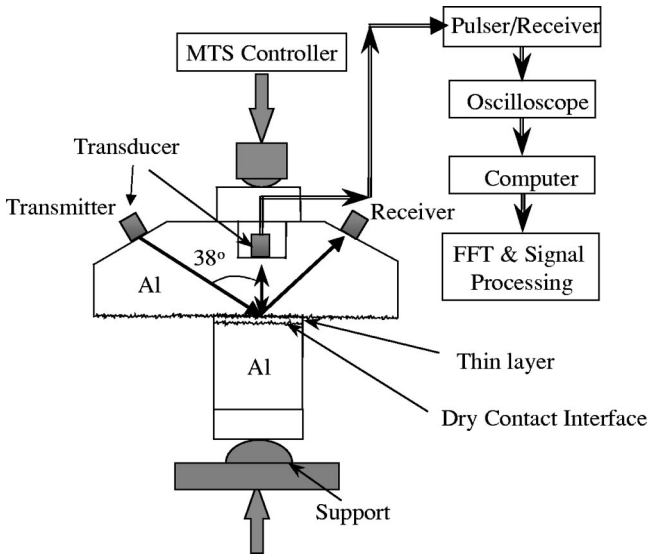


FIG. 6. Experimental setup for the determination of normal and oblique reflection signals from a thin layer embedded between two aluminum blocks under controlled load.

III. EXPERIMENT

The key element of this study is to validate experimentally the reconstruction algorithm and to demonstrate the separation of the bulk layer bulk properties from those of the interface. To do this we need the ability to measure the properties by independent techniques, which is a difficult task. We partially solved this problem by specially designing the experiments and using thin layers with independently measured properties and by obtaining controllable interfacial conditions between the layer and the substrates.

A. Thin layer in dry contact between substrates

1. Experimental procedure

In one series of experiments the measurements were performed on a sample with a thin aluminum layer positioned in dry contact between two aluminum substrates. The interfacial contacts are controlled by applying predetermined pressure on the layered system. The bulk properties of the thin layer are measured independently; they do not change under pressure, but the interfacial contacts vary significantly.[21–23] Our objective is to reconstruct these bulk and interfacial properties from ultrasonic measurements and to compare the reconstructed bulk properties with the original ones. The result of this comparison serves as a measure of the quality of the reconstruction of the properties including those of the interfaces.

The experimental setup is depicted in Fig. 6; this includes the mechanical system with load control, ultrasonic pulser/receiver and transducers, and signal acquisition system. The system is self-aligned, allowing homogeneous loading distribution at the contacting area. It is designed to perform simultaneous normal and angular acquisition of reflected signals as the system is subjected to increasing or decreasing controlled levels of loading.

In the experiments, a 0.77-mm-thick round plate with 25.4 mm diameter is mechanically loaded between the upper aluminum block A and the lower aluminum cylinder B. The nominal contact area and load level produced a range of applied pressure between 0–46 MPa. The contacting surfaces of the thin plates as well as aluminum substrates (blocks A and B) were roughened with 600×-grit sandpaper. Examining the surfaces after roughening showed a random distribution of asperities with an rms roughness of about 0.3 microns.

Three longitudinal wave transducers were positioned in the upper aluminum block: two for oblique incidence in pitch–catch mode and one for normal incidence in echo–pulse, as can be seen in Fig. 6. Transducers were all selected as broadband with central frequency 20 MHz; this allowed the presence of more than one resonance minimum in our experiments. The external loading was controlled via a Material Test System model 810, and the system operated in the range of applied load 0–10 klb (0–44.5 kN).

A critical part of the experiment is careful alignment of the contacting surfaces to obtain parallelism of the surfaces in contact; otherwise, nonuniformity of the interfacial stiffness along the contact area might take place (the misalignment effects have been discussed in detail in Ref. 27). To facilitate alignment, the lower part of the system (Fig. 6) has a hemispherical support; this allows two rotational degrees of freedom while providing normal force transfer.

Ultrasonic reflected signals were captured and digitized for further analysis as a function of the load cycle. The reflected spectra from the interface at each loading level were deconvolved with the reflected spectrum from the free surface.

2. Reflectivity spectra

The use of a thin aluminum layer between aluminum substrates allows isolating the effect of interfacial stiffness on reflected ultrasonic signals by eliminating the impedance mismatch between the layer and the substrates. When the thin plate and substrates have the same properties, the position and amplitude of the reflection minima will depend only on the interface stiffness, which is allowed to change from zero, at complete disbond, to infinity for perfect interface conditions; in this last case the reflection is zero. Lavrentyev and Rokhlin[21] investigated the application of the spring model to model the reflected signature from normal incidence of longitudinal waves, and established that the frequency minima shift can be used to estimate the normal interfacial stiffness.

Experimental results for the aluminum/aluminum layer/aluminum system using longitudinal waves at normal and 38 deg oblique incidence versus pressure on the sample are given in Fig. 7. Results are for nominal pressures ranging from 10 to 24 MPa with increments of 2 MPa. The reflected spectra from normal as well as oblique incidence shift to higher frequencies with increase of applied pressure (the resonance minima positions and its depth clearly identify the shift of the spectrum). The frequency minimum shift is about 0.4 MHz between 10 and 24 MPa. The shift of the frequency minima on pressure is similar in all pressure ranges, indicating that the saturation point (frequency independence) was not reached.

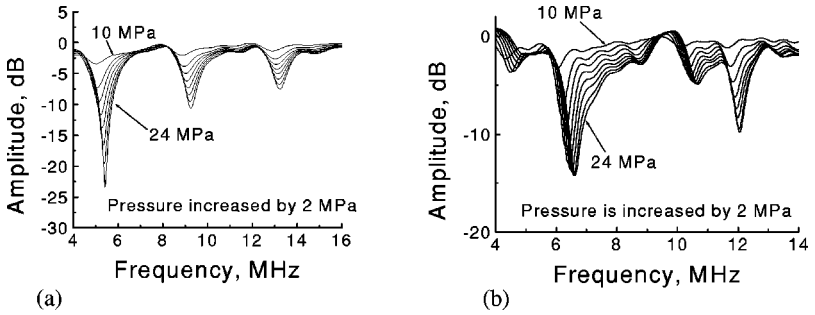


FIG. 7. Experimental reflection spectra (amplitudes versus frequency) for an aluminum/aluminum/aluminum system under different pressure: (a) normal and (b) angular incidence reflected spectra (38-deg incident angle).

The experimental results for oblique incidence of reflected longitudinal waves are given in Fig. 7(b). Again, the spectra of the reflected signal shift to higher frequencies with increase of pressure. The resonance minima at around 6 and 12 MHz are clearly shifted together with the local minima which appeared at approximately 9 and 10 MHz. The sensitivity of the minimum shift to pressure is slightly higher at low frequencies than at high frequencies. Similar results were obtained for a thin brass foil between Al substrates.

B. Environmentally degraded adhesive joint

The second set of validation experiments was performed on adhesive bonds with environmentally degraded interfacial properties. Adhesive joints were prepared with modified interfacial properties and nominally constant bulk properties of the adhesive layer. The reconstruction of properties was performed from experimental data using the inverse model, which includes transverse springs simulating poor contact between substrates and an adhesive layer. The experimental results were obtained for aluminum alloy bonded with commercial FM-73 adhesive. Interfacial degradation was controlled by time in the environment of NaCl saturated solution at 68 °C. The sample preparation process was discussed in detail in Refs. 26 and 28. To isolate the interfacial degradation from changes of bulk properties of the adhesive, aluminum adherents were prepared without anodization and priming, assuring fast interface degradation. In this case 100% interface degradation occurs relatively fast (in a few days). Ultrasonic data were collected using the dual-channel angle-beam contact scanning system described in Ref. 28. The normal and oblique reflected spectra from the adhesive bond were deconvolved with the reference signal as described in Refs. 25 and 28. We continue the degradation process after the measurements until the adhesive layer is easily separable from the substrates.

At this stage the adhesive layer is removed from the joint and its bulk properties are measured again (on a single layer) to compare with the original ones which were measured after bonding but before degradation, using the reconstruction method.[24]

The adhesive joint samples were exposed to the environment until moisture penetration was detected at one of the interfaces by the 50-MHz focused transducer C-scan in the reflection mode (at this frequency, the reflected signals from the bottom and top of the adhesive layer are separated and the transducer is focused on the top interface). Figure 8(a) shows a typical C-scan image obtained from a sample environmentally degraded for 73 h where the darker areas are associated with water accumulation at the interfaces (low reflection amplitudes indicating no disbond are bright areas). The image shows that the water-filled delamination is propagating inward from the edges. It is important to note the C-scan image is related to one interface. We have performed C-scans from both sides of the sample and found that the C-scan images of both interfaces are statistically similar for the type of environmental degradation performed in this

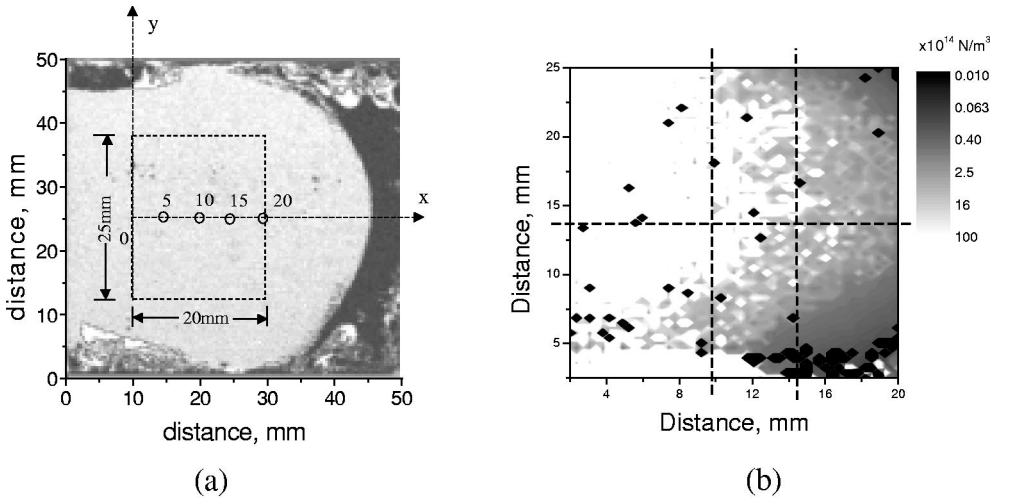


FIG. 8. (a) C-scan of damaged area with 50-MHz focused transducer; (b) reconstructed properties of transverse interfacial stiffness K_T' in the area marked by dashed box in (a).

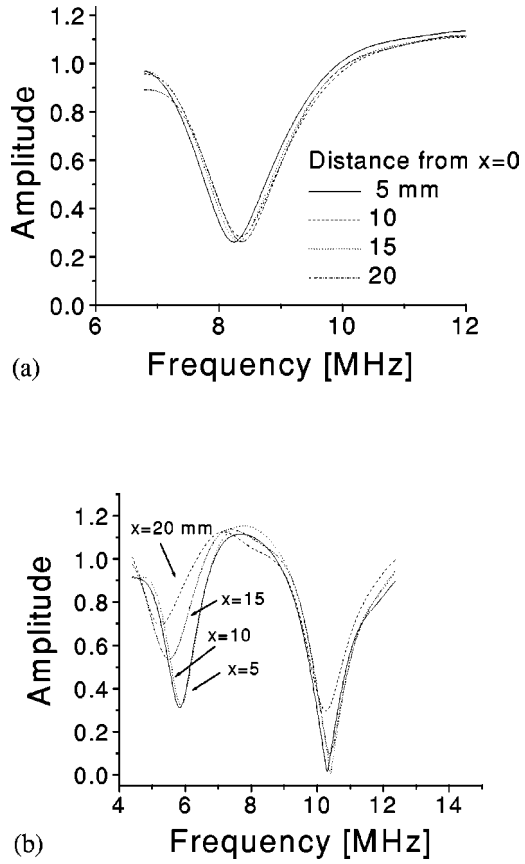


FIG. 9. Spectra of the reflected signal at different positions in the environmentally degraded area of the adhesive bond. (a) Normal incidence; (b) oblique incidence [the curve labels are as in (a)].

work. The 20- by 25-mm area, marked off by the dashed line in Fig. 8(a), exhibits no visible damage in the C-scan image. It was used for the dual beam (normal and oblique) ultrasonic scan and property reconstruction. An image of the reconstructed transverse interfacial spring K_T in this region is shown in Fig. 8(b). Due to reverberation in the adhesive of the signal used for reconstruction, the K_T image depends on the reflection from both interfaces as was discussed in Sec. II B.

As an example, Fig. 9 shows normal and oblique incidence reflection spectra for several points along the centerline x of the C-scan image in Fig. 9(a). The spectrum for normal incidence [Fig. 9(a)] at these points is practically unaffected until $x=20$ mm, which is close to the disbonded area. The normal spectrum, however, changes as the scan approaches the edge of the samples toward areas with complete disbonding. The obliquely reflected SV-wave spectra show a significant shift [Fig. 9(b)] as one approaches the edge of the scanned area. Therefore, the sensitivity of the obliquely incident waves to the onset of such damage at the interphase is stronger. This is in line with the model proposed by Lavrentyev and Rokhlin[20] to characterize adhesive bond degradation due to moisture migration to the interface. It accounts for formation of very thin water-filled interfacial microdisbonds that produce very little effect on longitudinal wave reflection at normal incidence. Such a hypothesis is supported by the infrared measurements performed at NIST,[29] which show that the thickness of the typical water layer diffused to the adhesive–adherent interface is below 40 nm. Such microdisbonds can be modeled by slip boundary conditions (zero stiffness transverse springs); however, the bonded area contains arrays of small microdisbonds which are modeled by finite transverse springs, which depend on the total microdisbonded area.[20]

IV. INTERFACE AND LAYER PROPERTY RECONSTRUCTION

The inverse algorithm was applied to reconstruct the set of parameters (3) and (4) from the normal and oblique experimental spectra. For the Al/Al/Al system the imaginary part of the complex interfacial spring corresponds to energy loss due to ultrasonic scattering at the contact roughness on the interfaces, and needs to be determined as a function of the nominal pressure together with other parameters. Attenuation in the aluminum layer is comparably low for both longitudinal and transverse waves and is set to be zero during reconstruction. As an example, Fig. 10 shows typical experimental (points) and simulated (lines) spectra after optimized matching of the spring model with the experimental spectra. Solid lines represent the spectrum calculated from the reconstructed layer and interface properties, and the points are the experimental data. The deviation between the experimental and calculated spectra serves as an indication of the quality of the reconstructed data. Both normal and oblique experimental spectra are well represented by the calculated spectrum for all features, including the local minima in the oblique spectra.

A summary of the reconstruction results of the bulk and the interfacial properties versus applied nominal pressure is given in Fig. 11. The results show expected trends; for example, Fig. 11(a) demonstrates the increase of interfacial normal and tangential springs with increasing pressure; however, the damping shown in Fig. 11(b) decreases rapidly with

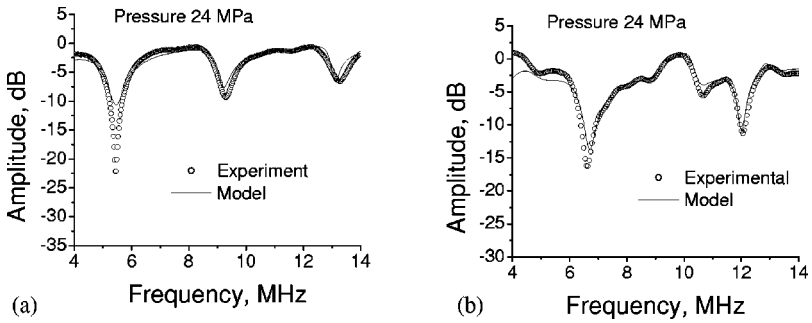


FIG. 10. Comparison between experimental and calculated reflection spectra for the aluminum/aluminum/aluminum system. Calculations (solid lines) were performed using reconstructed spring constants: (a) normal incidence, and (b) oblique incidence (38-deg incident angle).

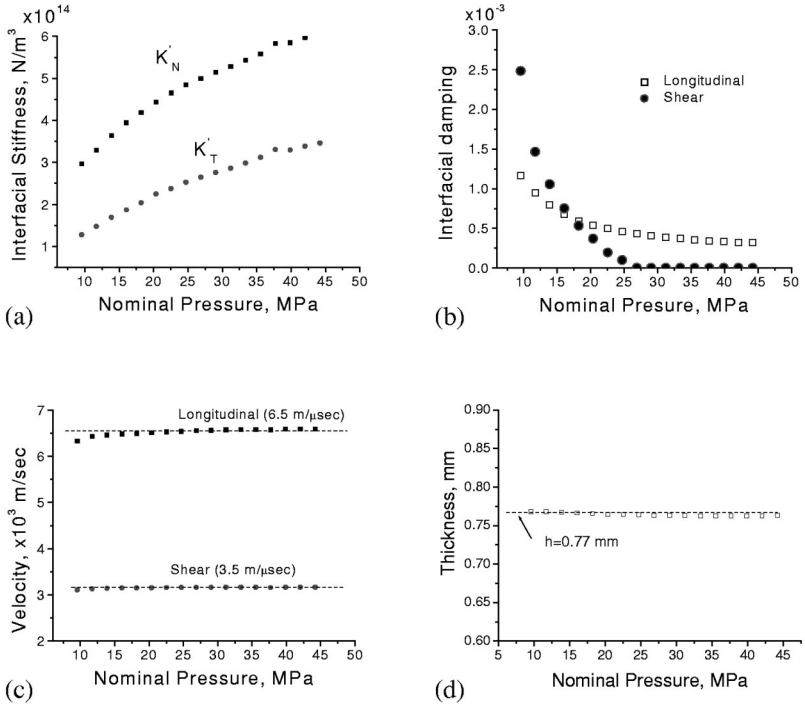


FIG. 11. Reconstructed properties for an Al layer between Al substrates under dry contact as function of the applied nominal pressure: (a) interfacial stiffness K_N', K_T'; (b) imaginary parts of the normal and transverse interfacial springs (damping); (c) longitudinal and shear velocities; and (d) thickness of the aluminum layer obtained. Dashed lines in (c) and (d) are properties determined on this thin aluminum layer by independent direct measurements.

pressure. This is due to reduction of the interface scattering with flattening of the asperities with pressure. The ratio K_T'/K_N' for the data in Fig. 11(a) has a value between 0.42–0.56. These values are reasonably close to those obtained in Refs. 9 and 23 (for extensive discussion of this ratio see Ref. 23). Reconstructed aluminum foil properties—velocity and thickness—do not change with the pressure. We performed direct measurements of the film thickness and shear and longitudinal velocities. Direct measurements are indicated in Figs. 11(c) and (d) by dashed lines. They are in good agreement with the reconstructed properties shown by solid and open points.

The bulk properties of the adhesive layer and the interfacial properties between the layer and the substrates were reconstructed from the experimental normal and oblique reflection spectra in the bonded area 20 by 25 mm marked by the dashed box in Fig. 8(a). As discussed in Sec. III B, the normal reflectivity spectra Fig. 9(a) are not sensitive to small environmental degradation, while the minima position for the angle beam spectra [Fig. 9(b)] exhibit significant shift. This is consistent with the model of water-filled microdisbonds at the interface[20] (tens of nanometers thick disbonds). We take in the reconstruction the normal spring $K_N=\infty$ (continuity of normal stresses and displacements) while keeping the K_T spring as unknown. There are two reasons for this decision: (1) Due to low sensitivity of the experimental normal spectrum minimum to degradation, the reconstruction of K_N' from these data is very susceptible to data scatter. (2) Due to limited transducer bandwidth, we can easily observe experimentally two minima for the oblique incidence in the reflection spectra at adhesive bond and only one minimum for the normal incidence [Figs. 9(a) and (b)]. As it was discussed in Sec. II B, from only one spectral minimum we cannot separate the longitudinal modulus of the adhesive from the normal K_N interfacial spring.

As discussed in Ref. 20, we can neglect viscous losses due to water accumulation at the interface. Also, the interfacial microdisbonds are very small and flat; at these conditions, in analogy with the results in Fig. 11(b) for the interfacial damping on a rough interface at higher pressure, it is reasonable to assume that the interfacial scattering and absorption are very small and therefore we take the imaginary part of the interfacial spring $K_T''=0$. Attenuation in the adhesive changes with degradation due to debonding of the knit carrier fibers in the adhesive and ultrasonic scattering.[30] Therefore, the attenuation for shear and longitudinal waves in the adhesive is assumed unknown and to be reconstructed. Experimentally we observe two easily measurable resonance minima at oblique incidence [Fig. 9(b)]; therefore, the K_T effect is separable from the adhesive bulk properties, and the interfacial transverse spring is reconstructed from the oblique spectrum.

An image of a reconstructed transverse interfacial spring K_T' is shown in Fig. 8(b). High values of K_T' are shown by white and low values by dark gray levels. Significant inhomogeneity is observed in the image with transition from high to low spring area. This contrasts with the C-scan image Fig. 8(a), which exhibits no changes in the same area. The transition zone in Fig. 8(b) appears ''mushy'' with highly variable transverse springs; this supports the water-filled microdisbond model[20] (black points are noise). The ultrasonic beam has about 5 mm diameter and thus averaged properties are measured in this area.

The reconstructed data at the centerline (at $y=0$) in the reconstruction area are shown in Fig. 12. The bulk properties (bulk and shear modulus, longitudinal and shear ultrasonic attenuation, density and thickness) did not show change with distance. This is expected since the samples stayed in the environment for only a few days. The interfacial transverse

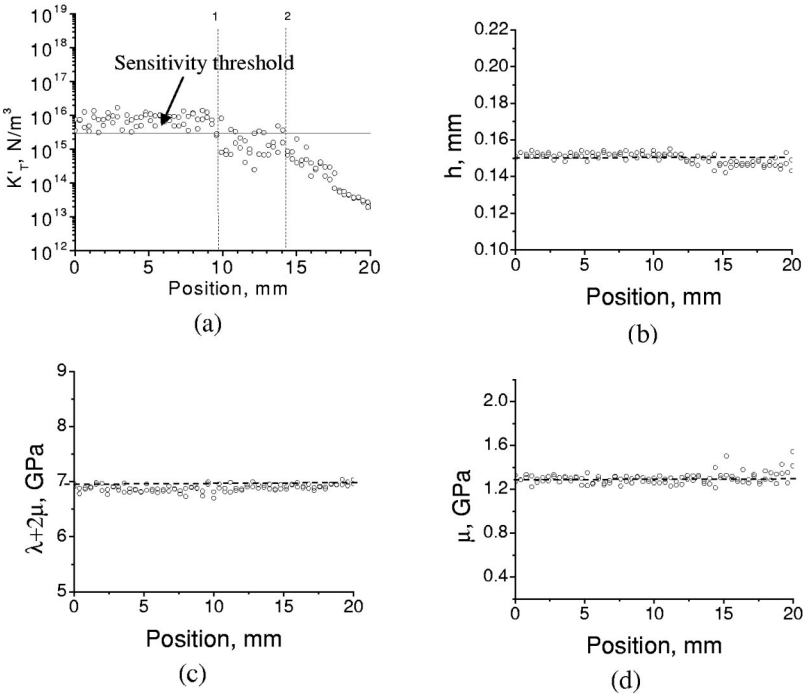


FIG. 12. Scan line of reconstructed properties of environmentally degraded adhesive bond: (a) transverse interfacial stiffness [horizontal solid line indicates the upper threshold level of the sensitivity to the K_T' reconstruction; the region from the left of the "1" vertical dashed line correspond to the "good bond," region between the lines "1" and "2" corresponds to the mushy area in Fig. 8(b)]; (b) thickness; (c) longitudinal modulus $(\lambda+2\mu)$; (d) shear modulus (μ) of the adhesive layer. The horizontal dashed lines in (b), (c), and (d) indicate the trend of properties variation with distance.

stiffness K_T' [Fig. 12(a)] shows a significant decrease as the position approaches the right edge. Note that in this range K_T' has a similar order of magnitude as that for the dry contact interface [Fig. 11(a)]. In the line scan of the reconstructed K_T' shown in Fig. 12(a), we should distinguish three regions: (1) The left region from the vertical dashed line "1" corresponds to the "good bond" (in the good bond area we cannot reconstruct the interface anomalies by this method in the 5–15-MHz frequency range). In this range the value of the spring is between the threshold level of the spring reconstruction sensitivity and $K_T'=\infty$, which corresponds to the perfect (welded) boundary conditions between the adhesive and the substrates. The threshold level ($K_T'\approx 3*10^{15}$ N/m^3) was assessed by the sensitivity simulation [see also Ref. 20 and Fig. 2(b)]. At K_T' above the threshold level the ultrasonic reflection signature does not depend on the spring value. At K_T' below the threshold level, frequency minimum shift is detectable with its decrease. Therefore, we depicted the reconstructed values of the good bond spring constant to be at this upper threshold level of the sensitivity (this shows some continuity of the experimental data; however, it does not imply that the actual spring constants are measured and have these values). (2) The middle transition region (between the vertical lines) corresponds to the mushy area in Fig. 8(b); it is smoothed by the finite size of the transducer beam (unsharpness effect). (3) From the right of the vertical dashed line "2" the bond is degraded.

V. SUMMARY AND CONCLUSIONS

A method for determination of the set of acoustical, geometrical, and interfacial properties of an isotropic layer embedded between two known substrates was developed. The properties include the layer density, elastic moduli, thickness, and interfacial normal and transverse spring constants. Also, either longitudinal and shear wave attenuations in the layer or the imaginary parts of the interfacial springs, which account for the wave scattering at the interface, are determined. In the reconstruction algorithm, the properties are combined in the form of eight nondimensional parameters, which are determined from reflection spectra at two incident angles: normal and oblique.

It was shown that the layer bulk and interface properties could be separated if several resonance minima in the reflection spectra are measured. If only one of such minima is recorded in both normal and oblique spectra, only effective properties of the layer can be reconstructed (those include combinations of layer bulk and interfacial properties). It was found that the attenuation in the layer and imaginary parts of the spring constants cannot be decoupled. It is suggested to select only two attenuation variables as unknown while the other two should be considered constant during reconstruction.

The algorithm was validated by specially designed experiments where layer properties were measured independently for comparison with the reconstructed values. One of the experiments is done on an aluminum foil in dry contact between two aluminum substrates. The attenuation in the foil was taken to be zero while the imaginary parts in the interfacial springs were considered unknown. All eight parameters were reconstructed. The reconstructed layer properties were found to be in good agreement with Al foil properties. Also, the interfacial spring ratio K_T'/K_N' was found to be close to that in the literature.[9,23] The second experiment was for environmentally degraded adhesive bonds prepared such that mainly interface properties were modified. The longitudinal spectra were found to be weakly dependent on measured position in the slightly damaged bonded area, while the oblique spectra varied significantly. Thus, for reconstruction the longitudinal spring was set infinite and K_T was assumed to be unknown and to be determined from experiment. For this case the attenuation in the adhesive layer was assumed

dominant and unknown, while the imaginary part of the K_T spring constant was taken to be zero. The reconstruction results for the adhesive show good agreement with independently measured layer properties.

ACKNOWLEDGMENTS

This work was partially sponsored by the Federal Aviation Administration under Contract No. DTFA03-98-D-00008, Delivery Order No. lA021 and performed by The Nondestructive Evaluation Program at The Ohio State University as part of the Center for Aviation Systems Reliability program through the Airworthiness Assurance Center of Excellence.

[1] J.-M. Baik and R. B. Thompson, ''Ultrasonic scattering from imperfect interfaces: A quasi-static model,'' J. Nondestruct. Eval. **4**, 177–196 (1984).
[2] Y. C. Angel and J. D. Achenbach, ''Reflection and transmission of elastic waves by a periodic array of cracks,'' J. Appl. Mech. **52**, 33–41 (1985).
[3] F. J. Margetan, R. B. Thompson, and T. A. Gray, ''Interfacial spring model for ultrasonic interactions with imperfect interfaces: Theory of oblique incidence and application to diffusion-bonded butt joints,'' J. Nondestruct. Eval. **7**, 131–152 (1988).
[4] F. J. Margetan, R. B. Thompson, J. H. Rose, and T. A. Gray, ''The interaction of ultrasound with imperfect interfaces: Experimental studies of model structures,'' J. Nondestruct. Eval. **11**, 109–126 (1992).
[5] K. Kendall and D. Tabor, ''An ultrasonic study of the area of contact between stationary and sliding surfaces,'' Proc. R. Soc. London, Ser. A **323**, 321–340 (1971).
[6] W. Wang and S. I. Rokhlin, ''Evaluation of interfacial properties in adhesive joints of aluminium alloys using angle-beam ultrasonic spectroscopy,'' J. Adhes. Sci. Technol. **5**(8), 647–666 (1991).
[7] S. I. Rokhlin and Y. J. Wang, ''Analysis of boundary conditions for elastic wave interaction with an interface between solids,'' J. Acoust. Soc. Am. **89**, 503–515 (1992).
[8] S. I. Rokhlin, M. Hefetz, and M. Rosen, ''An elastic interface wave guided by a thin film between two solids,'' J. Appl. Phys. **52**, 2847–2851 (1981).
[9] P. B. Nagy, ''Ultrasonic classification of imperfect interfaces,'' J. Nondestruct. Eval. **11**(3/4), 127–140 (1992).
[10] P. B. Nagy and L. Adler, ''Ultrasonic NDE of solid-state bonds: Inertial and friction welds,'' J. Nondestruct. Eval. **7**, 199–215 (1988).
[11] J. H. Rose, M. Bilgen, and P. B. Nagy, ''Acoustic double-reflection and transmission at rough water–solid interface,'' J. Acoust. Soc. Am. **95**, 3242–3251 (1994).
[12] A. B. Wooldridge, ''The effects of compressive stress on the ultrasonic response of steel–steel interfaces and of fatigue cracks,'' CEGB Report NW-SSD-RR-42-79, Berkeley Nuclear Laboratories (1979).
[13] N. F. Haines, ''The theory of sound transmission and reflection at contacting surfaces,'' CEGB Report RD-B-N4744, Berkeley Nuclear Laboratories, 1980.
[14] B. W. Drinkwater, R. S. Dwyer-Joyce, and P. Cawley, ''A study of the interaction between ultrasound and partially contacting solid–solid interface,'' Proc. R. Soc. London, Ser. A **452**, 2613–2628 (1996).
[15] A. M. Quinn, B. W. Drinkwater, and R. S. Dwyer-Joyce, ''The measurements of contact pressure in machine elements using ultrasound,'' Ultrasonics **39**, 495–502 (2002).
[16] L. J. Pyrak-Nolte, L. Myer, and N. G. W. Cook, ''Transmission of seismic waves across single natural fractures,'' J. Geophys. Res. **95**, 8617–8630 (1990).
[17] S. I. Rokhlin and A. I. Lavrentyev, ''Ultrasonic characterization of thin surface and interphase layers,'' in *Sensing for Materials Characterization, Processing and Manufacturing*, edited by G. Birnbaum and B. A. Auld (ASNT, Columbus, OH, 1998), pp. 47–83.
[18] S. I. Rokhlin, M. Ganor, and A. D. Degtyar, ''Ultrasonic characterization of plasma spray coating,'' in *Review of Progress in QNDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. 16B, pp. 1585–1591.
[19] Z. Guo, J. D. Achenbach, A. Madan, K. Martin, and M. E. Graham, ''Modeling and acoustic microscopy measurements for evaluation of the adhesion between a film and a substrate,'' Thin Solid Films **394**, 189–201 (2001).
[20] A. I. Lavrentyev and S. I. Rokhlin, ''Models for ultrasonic characterization of environmental degradation of interfaces in adhesive joints,'' J. Appl. Phys. **76**, 4643–4650 (1994).
[21] A. I. Lavrentyev and S. I. Rokhlin, ''Ultrasonic spectroscopy of imperfect interfaces between a layer and two solids,'' J. Acoust. Soc. Am. **103**, 657–664 (1998).
[22] A. Baltazar, S. I. Rokhlin, and C. Pecorari, ''On the relationship between ultrasonic and micro-structural properties of imperfect interfaces in layered solids,'' in *Review of Progress in NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1999), Vol. 18B, pp. 1463–1470.
[23] A. Baltazar, S. I. Rokhlin, and C. Pecorari, ''On the relation between ultrasonic micromechanic properties of contacting rough surfaces,'' J. Mech. Phys. Solids **50**, 1397–1416 (2002).
[24] A. Lavrentyev and S. I. Rokhlin, ''Determination of elastic moduli, density, attenuation, and thickness of a layer using ultrasonic spectroscopy at two angles,'' J. Acoust. Soc. Am. **102**, 3467–3477 (1997).
[25] L. Wang, B. Xie, and S. I. Rokhlin, ''Determination of embedded layer properties using adaptive time-frequency domain analysis,'' J. Acoust. Soc. Am. **111**, 2644–2653 (2002).
[26] A. Baltazar, B. Xie, L. Wang, and S. I. Rokhlin, ''Ultrasonic determination of environmental degradation of interfacial properties in adhesives bonds,'' in *Review of Progress in QNDE*, edited by D. O. Thompson and D. E. Chimenti (AIP Vol. 615, Melville, NY, 2002), Vol. 21B, pp. 1157–1164.
[27] B. Drinkwater and P. Cawley, ''Measurement of the frequency dependence of the ultrasonic reflection coefficient from thin interface layers and partially contacting interfaces,'' Ultrasonics **35**, 479–488 (1997).
[28] S. I. Rokhlin, A. Baltazar, B. Xie, J. Chen, and R. Reuven, ''Method for monitoring environmental degradation of adhesive bonds,'' Mater. Eval. **60**(6), 795–801 (2002).
[29] T. Nguyen, E. Byrd, C. Lin, and Dale Bentz, ''A novel technique for measurements of water at the coating/metal interface,'' Adv. Composites Mater. **19**, 1051–1060 (1991).
[30] S. I. Rokhlin, A. Baltazar, B. Xie, and J. Chen, ''Ultrasonic characterization of accelerated environmental degradation of adhesively bonded joints,'' in *Adhesive Joints: Formation, Characteristics and Testing*, edited by K. L. Mittal (VSP, Utrecht, Boston, 2002), Vol. 2, pp. 337–360.

Radiated fields of capacitive micromachined ultrasonic transducers in air

David A. Hutchins, James S. McIntosh, Adrian Neild, and Duncan R. Billson
School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom

Russell A. Noble
QinetiQ (Malvern), St. Andrews Road, Great Malvern, Worcestershire WR14 3PS, United Kingdom

(Received 13 February 2003; revised 18 June 2003; accepted 1 July 2003)

This paper presents a study of the radiated ultrasonic fields of capacitive micromachined ultrasonic transducers (often referred to as cMUTs) in air. These fields were modeled theoretically and then compared to the experimental near-field amplitude variations and directivity patterns of square cMUTs. The good agreement between theory and experiment indicates that the devices can be approximated to plane piston radiators. The fields of multiple elements driven in phase on the same silicon substrate are presented, where again comparison is made to theory. The results indicate that individual elements are unaffected by radiation through the silicon substrate from adjacent devices. It will also be demonstrated that it is possible to use the knowledge of these fields to develop air-coupled ultrasonic surface imaging systems. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1604120]

PACS numbers: 43.35.Zc, 43.38.Bs [DEC]

I. INTRODUCTION

There has been rapid progress in recent years in the design, fabrication, and modeling of completely micromachined ultrasonic transducers (or MUTs), using thin membranes suspended above a cavity to give a capacitive device (a cMUT). There have been two main approaches to the fabrication and the resultant geometric characteristics of cMUTs. The approach adopted by many workers in this area[1,2] is that of a multicellular structure, fabricated from many smaller capacitive resonant cells. The cMUT consists of a series of small hexagonal cells, with typical lateral dimensions of between 25 and 100 μm, fabricated using a low-pressure chemical vapor deposition (LPCVD) process. Each individual cell contains a 3.5-μm-thick Si_3N_4 (silicon nitride) membrane, suspended over a 1-μm cavity, which is vacuum sealed during the manufacturing process. The result is a capacitive ultrasonic transducer which is able to provide ultrasonic output over a range of frequencies in the MHz region.[3] Other devices such as those produced by Eccardt[4–6] and Caliano *et al.*[7] follow similar design principles. The small size of each cell means that a working device is typically made up of many such cells, connected together to make a larger device. The frequency response of an individual cell typically depends on both the geometry of the cell and the physical membrane characteristics.[8]

If larger cells are used, such that the cell diameter is much greater than either the membrane thickness and/or the gap behind the membrane, then it is possible to obtain a device with a frequency response which, to a first approximation, is independent of the cell diameter. This means that the cMUT can be considered as a much closer approximation to a plane piston, surrounded by an infinite rigid baffle. The devices described by Noble *et al.*[9] are single cells with dimensions in the range 1 to 5 mm, and are used in the present work. A typical device is shown in Fig. 1. It is manufactured in a PECVD process, with 0.5–2-μm-thick Si_3N_4 membranes and air gaps of 1–2 μm. Note that the stress in the membrane is carefully controlled during manufacture, to give stable membranes over large device dimensions. Note also that a regular array of holes is visible in the top electrode. These are etch holes, required during manufacture for the removal of the sacrificial layer. These are left open for operation in air, but can be sealed if necessary for water immersion. In the work reported here, all measurements have been conducted in air for characterization purposes.

Previous work[10] has shown that such devices can be excited at ultrasonic frequencies, and that the displacement of the membrane is typically in phase across the membrane. A typical device has a response with a maximum output at around 1.2 MHz, this value being independent of size above dimensions of $\sim$0.25 mm. The wavelength (λ) at this frequency in air is $\approx$0.3 mm. Hence, the device widths are $\approx 3\lambda$. This will result in a complicated diffraction field, with a non-negligible near-field region. The exact nature of the radiated pressure in both the near-field and far-field regions is thus of interest.

II. THEORETICAL PREDICTION OF THE ACOUSTIC PRESSURE FIELD OF A CMUT IN AIR

Huygen's principle treats a finite source of sound as an assembly of point sources, and the pressure field can be modeled using a surface integral over the area of the source. There are several ways this problem can be tackled.[11] In this work, a convolution method, based on the Rayleigh surface integral, will be used, an approach sometimes called the impulse response method. The approach assumes that the source acts as a planar piston, where the velocity across the face of the source is constant, and that the rest of the plane

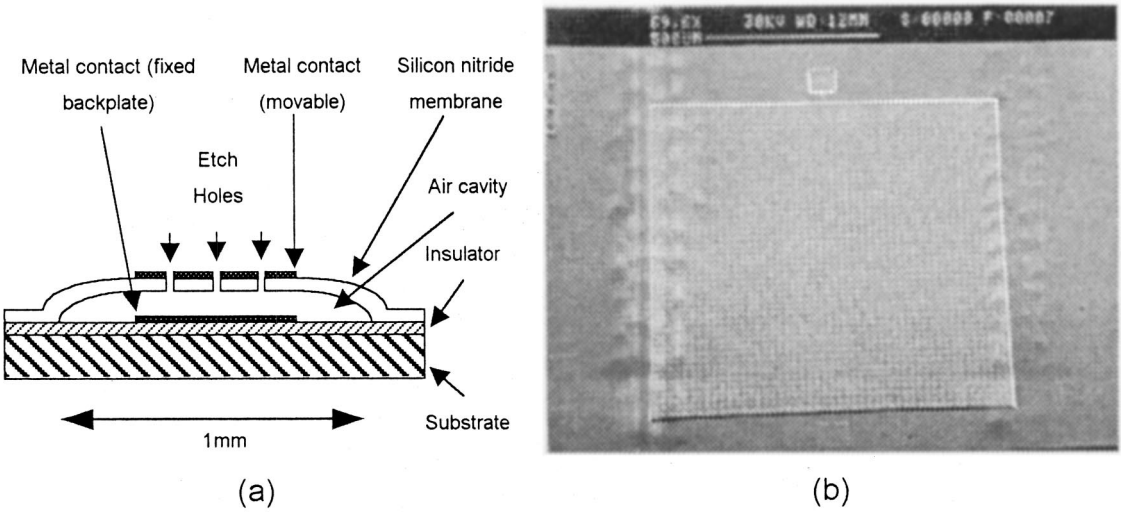


FIG. 1. (a) Schematic diagram of the cMUT design studied in this paper, and (b) a photograph of a 1-mm-square device.

containing the source, termed the baffle, is rigid (and hence the velocity is zero). Rayleigh stated that the time-dependent velocity potential, $\phi(M,t)$, is

$$\phi(M,t)=\int_s \frac{v(t-r'/c)}{2\pi r'}dS, \tag{1}$$

where $v(t)$ is the normal velocity across the area of the source within the plane in which it moves, M defines the point at which ϕ occurs at time t, and r' is the distance from the point M to the surface element δS (see Fig. 2).[12] The velocity potential is related to the pressure by

$$p(M,t)=\rho\frac{\partial\phi(M,t)}{\partial t}, \tag{2}$$

where ρ is the density of the acoustic transmission medium. By the use of the Dirac delta function and the convolution operator, the problem can be re-expressed in any of three ways

$$p(M,t)=\rho\frac{\partial[v(t)*h(M,t)]}{\partial t}, \tag{3}$$

where $v(t)$ is the velocity of the source, and $h(M,t)$ is the scalar impulse response generated by a Dirac delta function movement of the source, which is given by

$$h(M,t)=\int_s \frac{\delta(t-r'/c)}{2\pi r'}dS. \tag{4}$$

By examining Eq. (4) and Fig. 2, it can be seen that for a given point M, the area of the source which contributes to the scalar impulse response at a given time t is the area which lies at a distance of $r'=ct$ from the point. Within the arc, the Dirac delta function is unity (i.e., $\delta(t-r'/c)=1$); for all other points on the source, it is zero. This is referred to as the sifting function of the Dirac delta pulse.[13] In effect, Eq. (4) states that the scalar impulse response is the area of

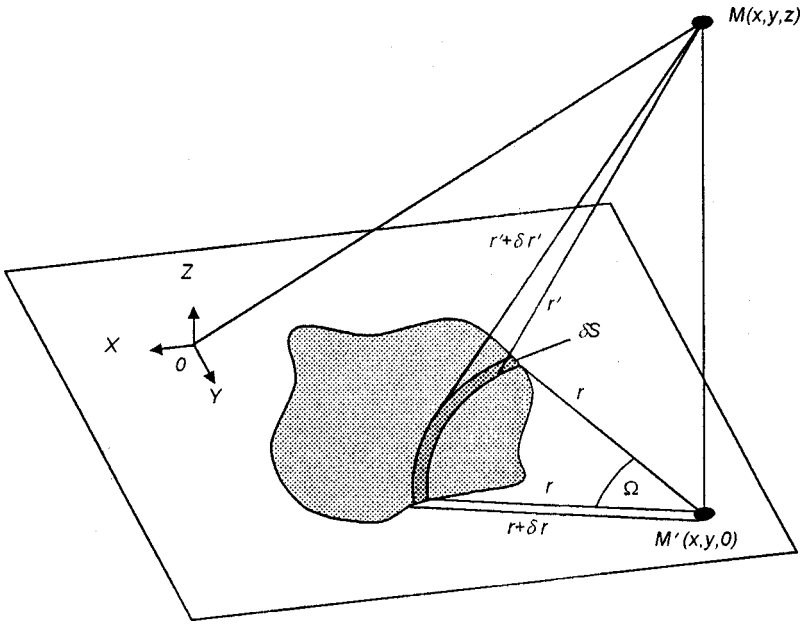


FIG. 2. Diagram showing the geometry used in the calculation of the field from an arbitrary plane piston in an infinite planar baffle.

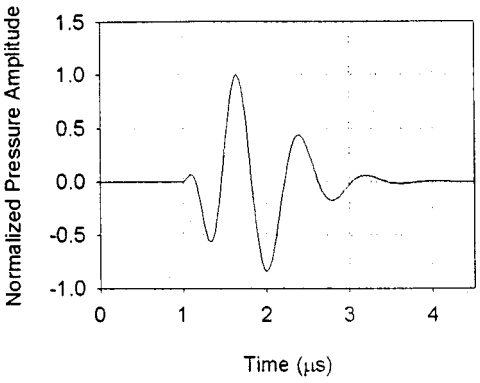


FIG. 3. Simulated piston velocity waveform, used to generate the predicted radiated fields of cMUTs.

the source at a distance of ct from the point M, divided by $2\pi r'$.

To simplify Eq. (4), consider the small area in the plane of the piston δS as r' is increased to $r'+\delta r'$, as shown in Fig. 2. This area can be found by considering the sphere centered at the point M. The plane of the piston will dissect this sphere, and the line of dissection will be a circle centered at M. If the coordinate system is set up as shown in Fig. 2, if the distance to point M is r' then the radius of this circle is $r=((r')^2-z^2)^{1/2}$, and similarly if the distance to point M is $r'+\delta r'$ then $r+\delta r=[(r'+\delta r')^2-z^2]^{1/2}$. This in turn means that $r\,\delta r=r'\,\delta r'$.

The small area, δS, considered in Fig. 2 can be expressed as $\Omega.r.\delta r$, which equates to $\Omega.r'.\delta r'$. Hence, the surface integral expressed in Eq. (4) can be expressed as an integral with respect to $\delta r'$ over the infinite range of r'

$$h(M,t)=\int_0^\infty \frac{\delta(t-r'/c)\Omega(M,r').r'}{2\pi r'}dr', \tag{5}$$

where the angle Ω has been expressed as a function of the field point M, and the distance from this point to the surface element r'. If $\tau=r'/c$, then

$$h(M,t)=\int_0^\infty \frac{\delta(t-\tau)\Omega(M,\tau).c}{2\pi}d\tau. \tag{6}$$

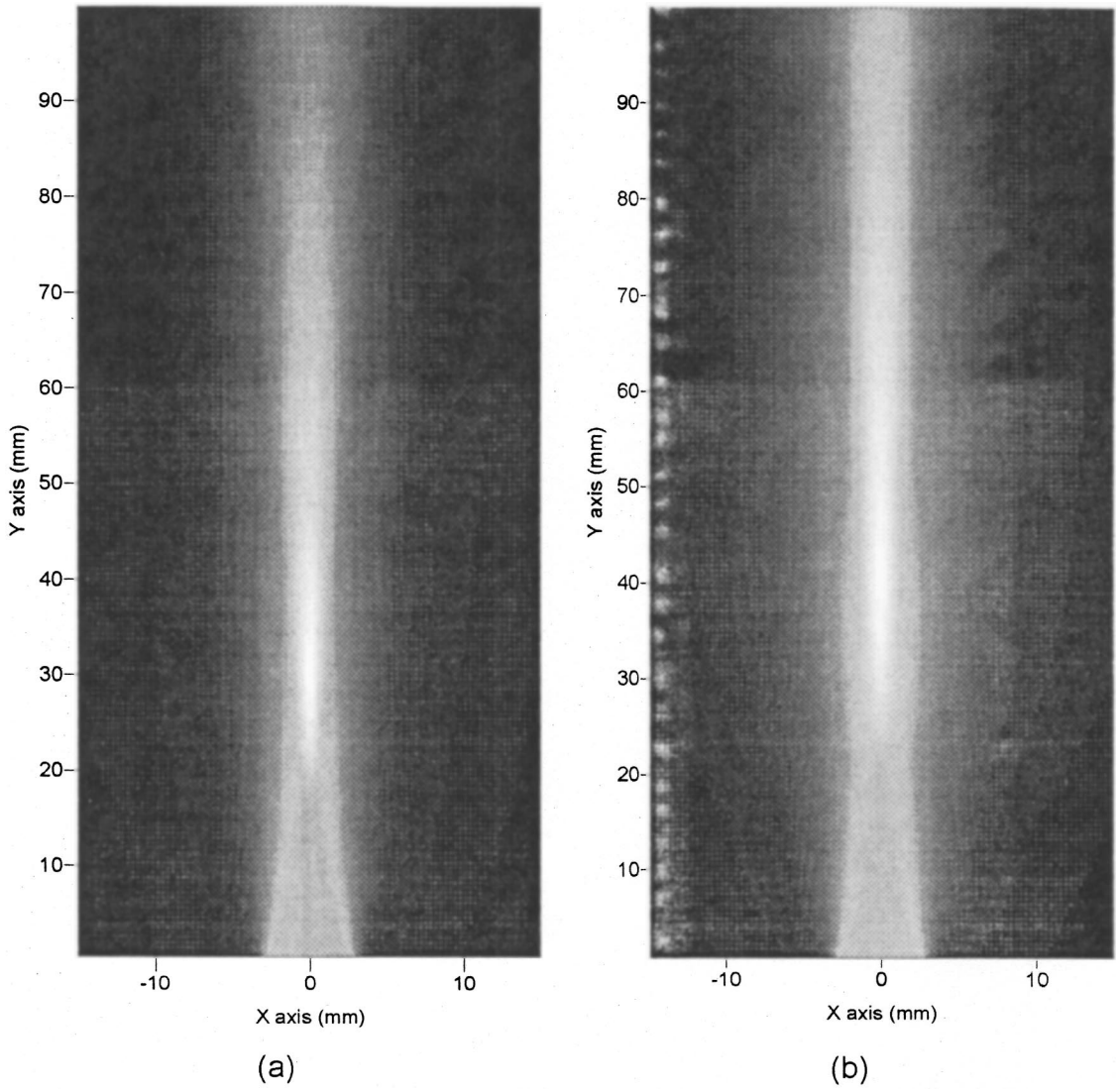


FIG. 4. Predicted radiated fields for (a) a 5-mm-diameter circular transducer, and (b) a square transducer of side 5 mm, using the piston response of Fig. 3.

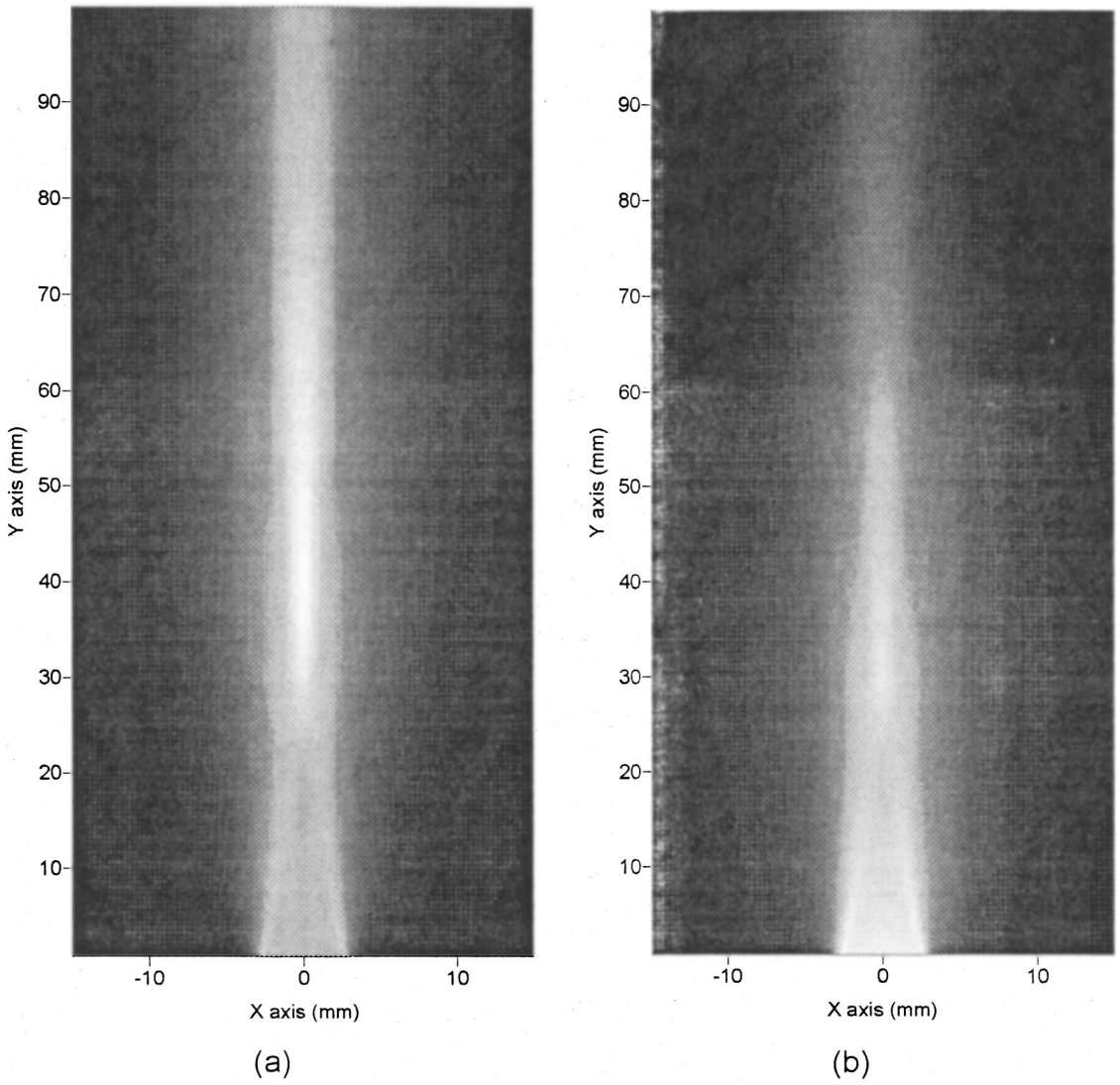


FIG. 5. Predicted radiated field for a 5-mm-square transducer after correction for attenuation in air via Eq. (8).

As discussed earlier, the sifting function of the Dirac delta is such that this integral can be simplified to

$$h(M,t)=c.\frac{\Omega(M,t)}{2\pi}, \tag{7}$$

where the angle $\Omega(M,t)$ refers to the angle subtended across the plane of the piston, from the point M at time t. This is a result which lends itself well to numerical modeling in which subsequent time steps are considered. In this way, the problem is reduced to one of computational geometry. The result is a time-limited function, where the limits are the shortest and longest path lengths from the point to the source.

The way the problem is approached has been described for an arbitrary piston shape. However, the implementation of this approach has been described in the literature for both circular[13,14] and rectangular pistons.[15,16] In this work, the solution suggested by San Emeterio will be used for the square sources.[15]

Figure 4 shows the predicted features of radiated pressure fields generated in air by circular and square plane piston sources. The prediction used a velocity waveform $[V_0(t)]$ in the form of a broad-bandwidth transient signal, similar to that observed experimentally for some cMUTs when driven with an impulse, an example of which is shown in Fig. 3. The radiated fields in Fig. 4 are shown for both a 5-mm-diameter circular transducer and a 5-mm-square transducer. These have been calculated using a MATLAB™ implementation,[17,18] and are based on the model proposed by Ullate and San Emeterio.[19,20] In all cases, the fields are plotted as spatial variations in peak-to-peak pressure amplitude. The radiated field in both cases can be explained by using the concept of plane waves generated from the face of the device, and edge waves from the transducer edges. Both near-field regions contain a triangular area of uniform amplitude starting at the transducer face, caused by a lack of interference between plane and edge waves. However, these contributions overlap at certain locations, causing the ridges that can be seen in both cases, starting at the transducer edges and converging towards the axial maximum. This axial maximum represents the position of the near-field/far-field boundary. Thereafter, the pressure field decays with distance, as expected in the far field. Note that there is a degree of similarity in the two fields. However, it can be seen that the square transducer has a slightly extended near field region,

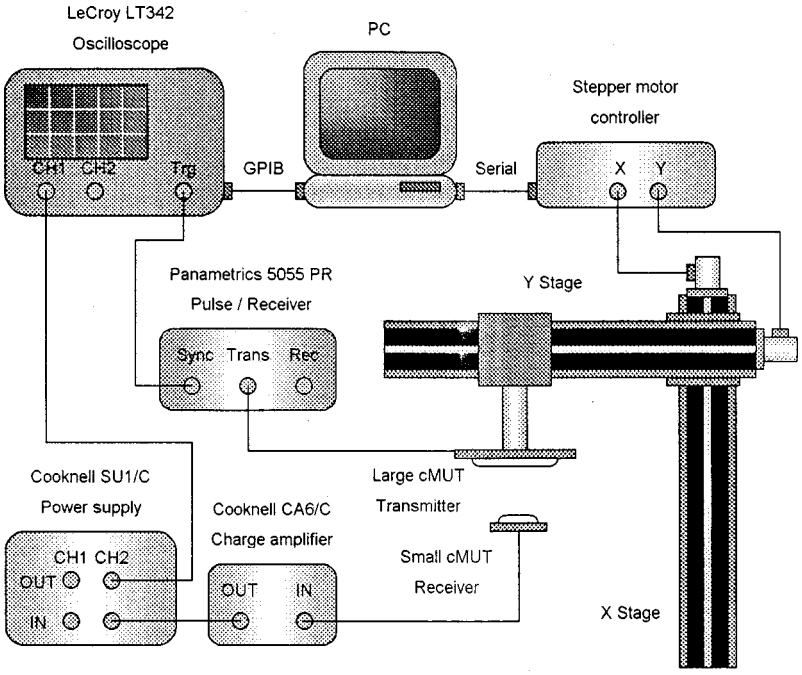


FIG. 6. Experimental apparatus for the mapping of the ultrasonic field radiated by a cMUT.

with the boundary to the far field being approximately 10 mm further out from the source. This might be expected from the greater area of the square source.

The above model predicts the radiated sound field from a piston in a nondissipative, isotropic, homogeneous medium. However, it is known that at ultrasonic frequencies in air, this assumption cannot be made. There have been several studies concerning this, and Bond *et al.*[21] have provided a simple expression for ultrasonic absorption in air at high frequencies. This has been applied to the fields predicted by the models in comparisons presented below, using the following expression:

$$\alpha_{cr}=\frac{15.895\times 10^{-11}\left(\frac{T}{T_o}\right)^{1/2} f^2}{\left(\frac{p}{p_o}\right)}, \tag{8}$$

where α_{cr} is the attenuation factor (dB/m), f is the frequency, T and p are the measured temperature and pressure, respectively, T_o is the reference temperature of 293.15 K, and p_o is the reference pressure of 101.325 kPa.

This expression holds for a temperature range of 0 °C–40 °C and for pressures below 200 kPa (2 atmospheres). Figure 5 shows the radiated field for the same 5-mm-square transducer as presented in Fig. 4, but with the effect of attenuation added via Eq. (8). It can be seen that there is a significant increase in axial amplitude decay with distance from the transducer face, as expected.

III. RADIATED FIELD MEASUREMENTS AND COMPARISON TO THEORY

A. Transient excitation

In order to compare experimental fields with predictions of the above model, the fields of several cMUTs have been mapped experimentally, and compared to the fields predicted theoretically. This has been achieved for square devices with a range of sizes, namely 1, 2, 3, and 5 mm.

Measurement of spatial variations in the radiated field required a miniature detector for use in air with sufficiently small lateral dimensions, when compared to the smallest cMUT to be mapped and the shortest wavelength of interest. It also had to have a suitable bandwidth and sufficient sensitivity in air to allow measurement of the received signals. For this purpose, a 500-μm-square cMUT was used, of the same basic design as the sources. As stated previously, the cMUT frequency response is independent of size for cMUTs with lateral dimensions greater than 0.25 mm. These smaller

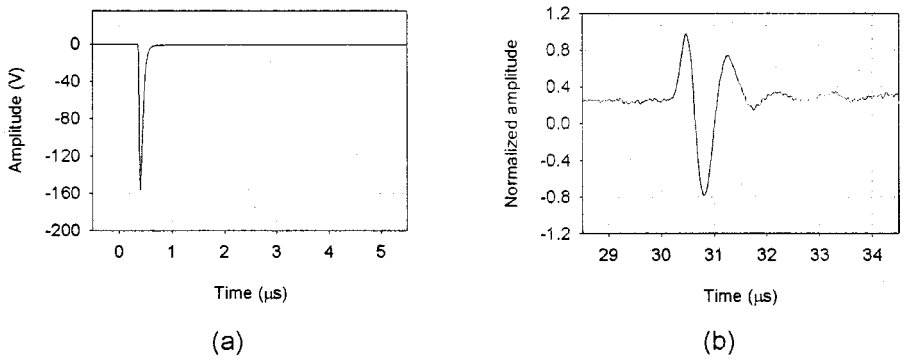


FIG. 7. (a) The impulsive drive signal generated by the Panametrics 5055PR pulser–receiver, and the (b) resultant ultrasonic signal generated in air by a 1-mm-square cMUT.

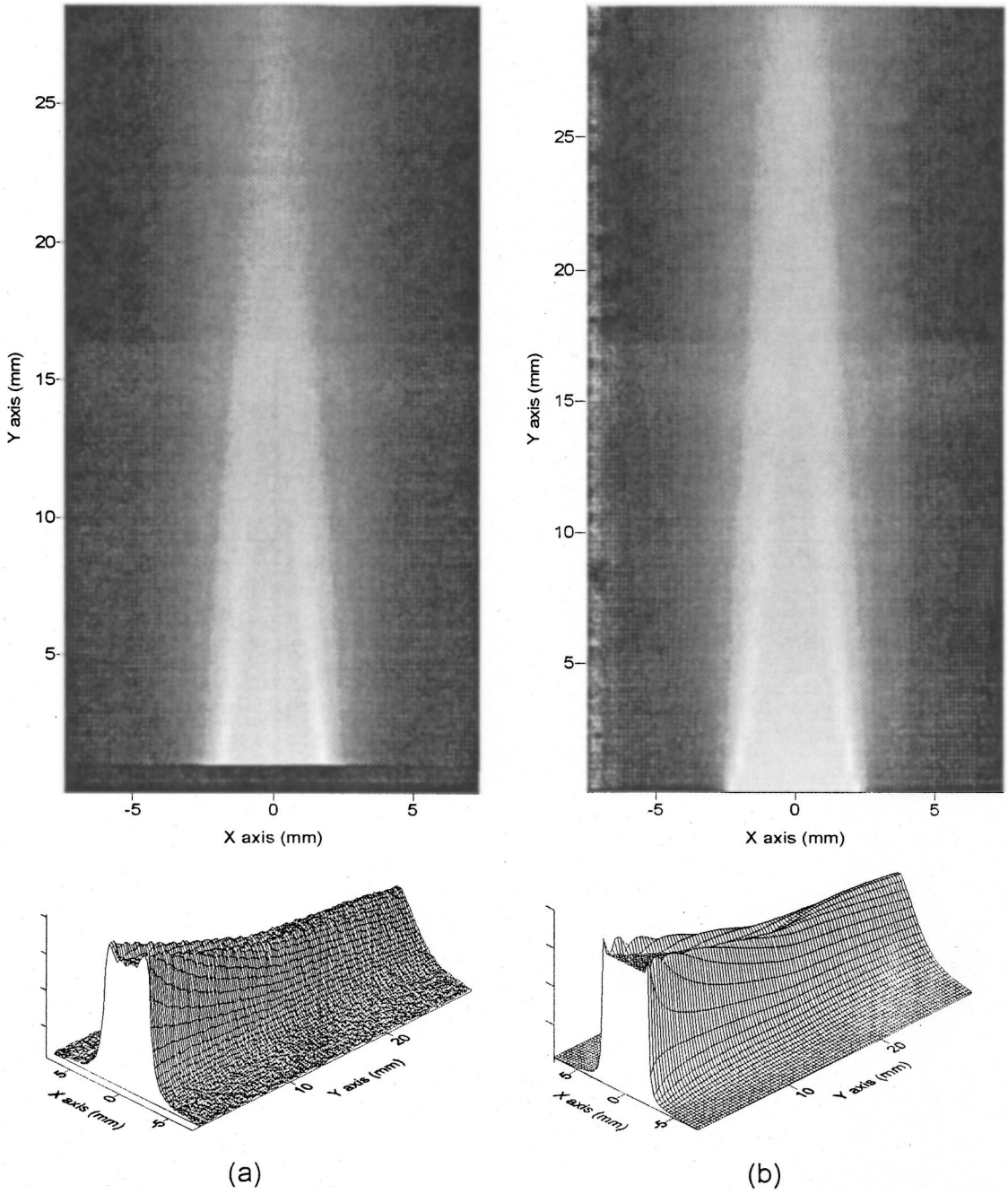


FIG. 8. Comparison of (a) experimental and (b) theoretical radiated fields for a 5-mm-square air–coupled cMUT.

cMUTs therefore have a suitable frequency response for mapping the field of the larger cMUT. When compared to alternative designs (e.g., a polymer-filmed capacitive device or a piezoelectric design with a matching layer) this was found to give excellent results. Note that the receiver was not calibrated in terms of sensitivity, and hence the resultant field measurement was a relative one. Hence, experimental values of peak-to-peak amplitude were normalized relative to the signal on-axis close to the transducer face. Note that cMUTs smaller than 0.5 mm could not be used as miniature detectors, as they did not have sufficient sensitivity for the detection of ultrasonic signals from the devices under test.

Figure 6 shows a schematic diagram of the apparatus used to scan experimental fields. The scanning was performed with the aid of a computer-controlled $X-Y$ stage, and custom software implemented in LABVIEW™. The source was driven by a 250-V transient voltage pulse from the Panametrics 5055PR pulser unit, with a rise time of $\sim$50 ns. This effectively simulated an inverted impulse (Dirac delta function) to cover the known frequency response of the source. Figure 7(a) shows the voltage drive signal used, and Fig. 7(b) shows a typical signal received on-axis. Although the generation signal from the pulser–receiver is close to an inverted Dirac delta function, the frequency response of the cMUT results in the output signal shown, with a peak response at approximately 1.2 MHz. Note that no dc bias was applied to the source, although receiver operation required a bias dc bias voltage of 30 V.

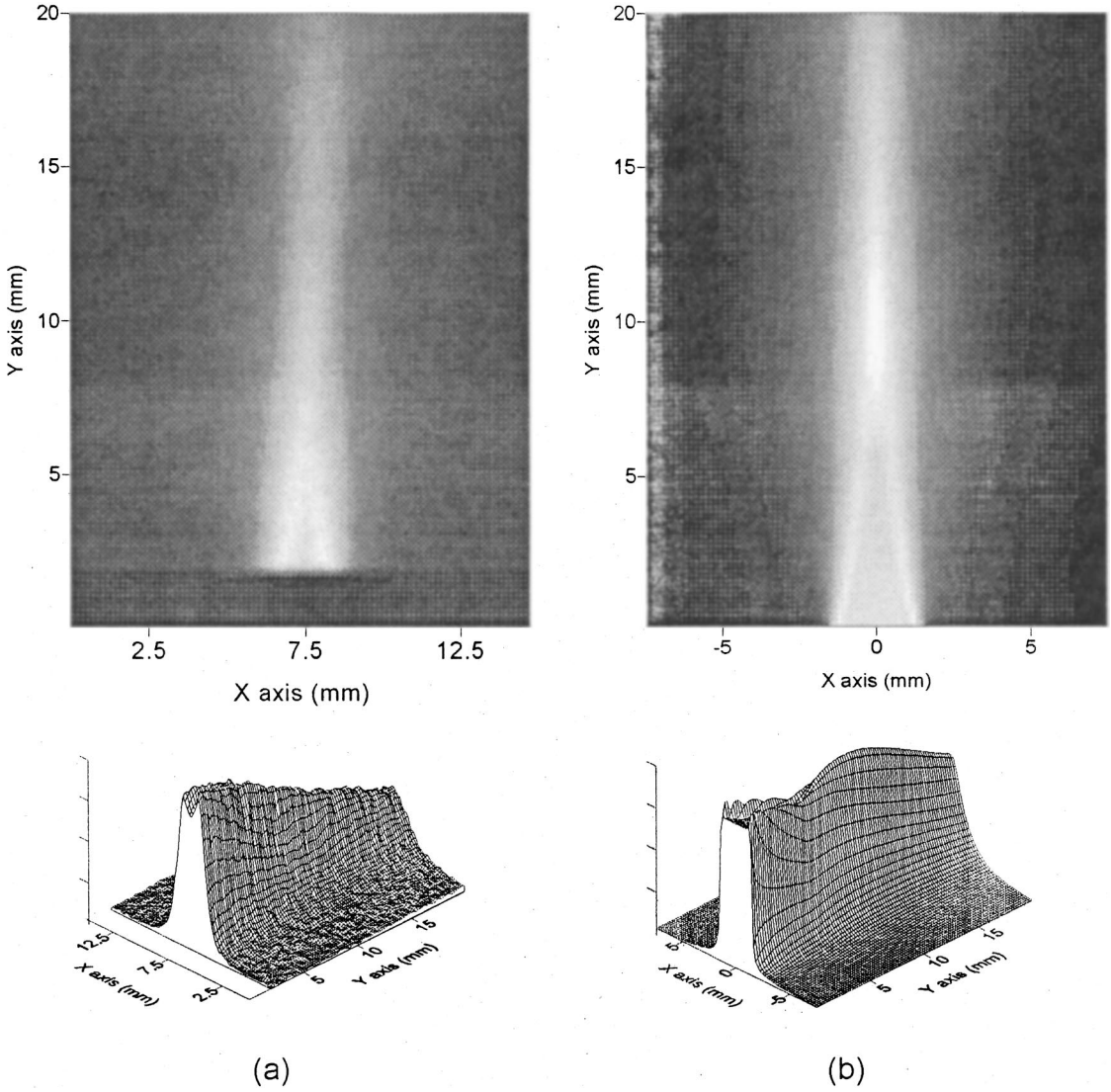


FIG. 9. Comparison of (a) experimental and (b) theoretical radiated fields for a 3-mm-square air–coupled cMUT.

For scanning purposes, the source was mounted on the $X-Y$ stage, and the stage was driven by a custom-designed motor controller, which was in turn controlled by a PC. The source was moved in a two-dimensional $X-Y$ plane in 1-mm steps, with the 500-μm cMUT receiver fixed in position. The receiver had a 30-V dc bias applied to it via the Cooknell CA6/C charge amplifier, the output from which was fed to a LeCroy LT342 oscilloscope under PC control. At each node of interest in the ultrasonic field, a time waveform was recorded by the PC for later analysis and field plotting. At the start of each scan, the source and receiver cMUTs were carefully aligned axially, and the scan initiated with reference to the axial start point, which was as close as practical to the surface of the cMUT. The limiting factor to this alignment was the cMUT packaging and bond wires. This meant that the probe could not be used to measure a radiated waveform arriving at an axial distance of less than 2 mm from the front of the cMUT for each set of scanned data. The recorded data were processed with custom signal-processing software, again implemented in LABVIEW™. In this instance, the software picked the maximum peak-to-peak amplitude in the waveform from each point in the radiated field, and placed it in a matrix. This could be read and output with the SURFER™ 3D imaging software.

The results obtained for the four sizes of square cMUT sources are shown in Figs. 8–11. The section of missing data at the front of each experimental field plot is due to the 2-mm clearance required for the scanned receiver mentioned above. In all four cases, a comparison is also presented to theoretical predictions for a square plane piston of the same dimensions. The field modeling software is designed to also output its results to a SURFER™-compatible file, thus ensuring that variations in output field patterns are only due to differences in field data. In all cases, there is very good agreement between theory and experiment. For the 5-mm source, Fig. 8, the extensive area of uniform amplitude at small axial distances can be seen, with the characteristic ridge converging towards the axial maximum. As the source dimensions decrease, the near-field/far-field boundary moves towards the transducer face, as is evident in both theory and experiment in Figs. 8–11.

There are small differences between the measured and the predicted data, and this is thought to be due to the theoretical assumption of a point receiver, whereas the receiver is

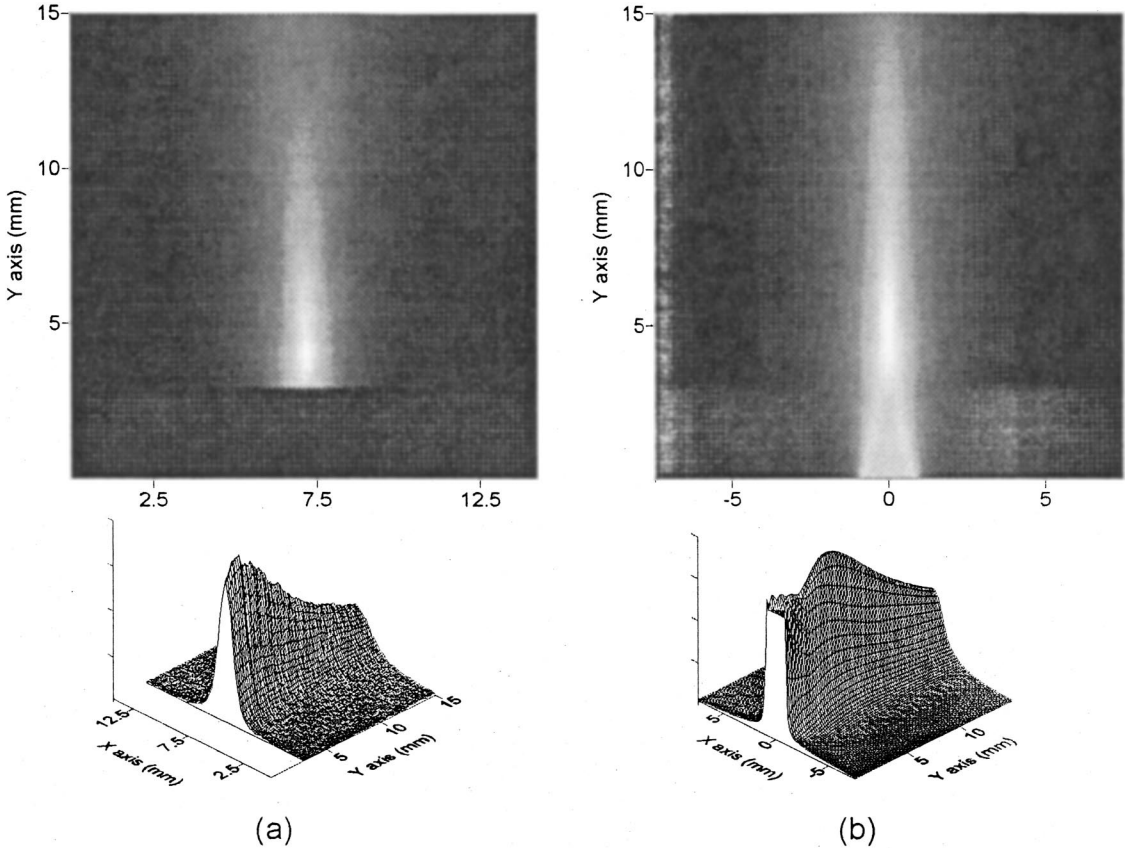


FIG. 10. Comparison of (a) experimental and (b) theoretical radiated fields for a 2-mm-square air–coupled cMUT.

in fact 500-μm wide. This is about 1.5λ (λ being the radiated wavelength) at the highest frequencies of interest. The receiver dimensions were therefore not negligible, and further work is currently being performed to predict the field that would be measured by a square receiver of finite size. Note that the effect of directivity on air-coupled transducer performance has been noted previously, and corrections applied.[22,23] However, it can be seen that the measured data

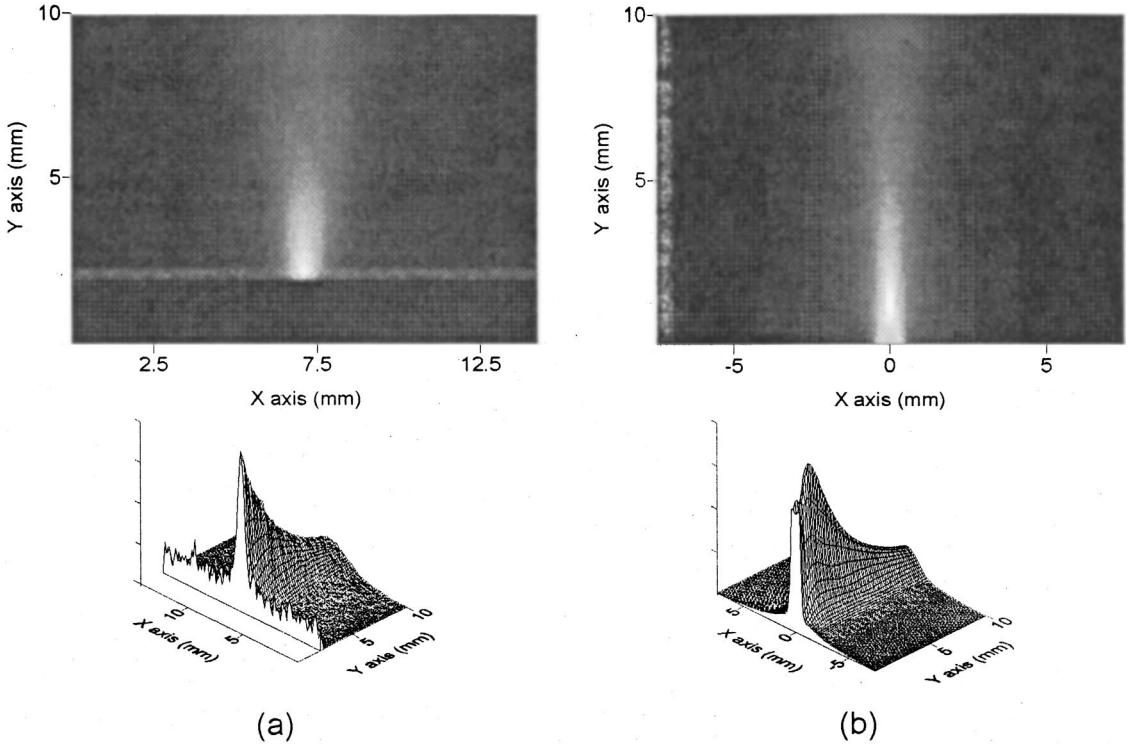


FIG. 11. Comparison of (a) experimental and (b) theoretical radiated fields for a 1-mm-square air–coupled cMUT.

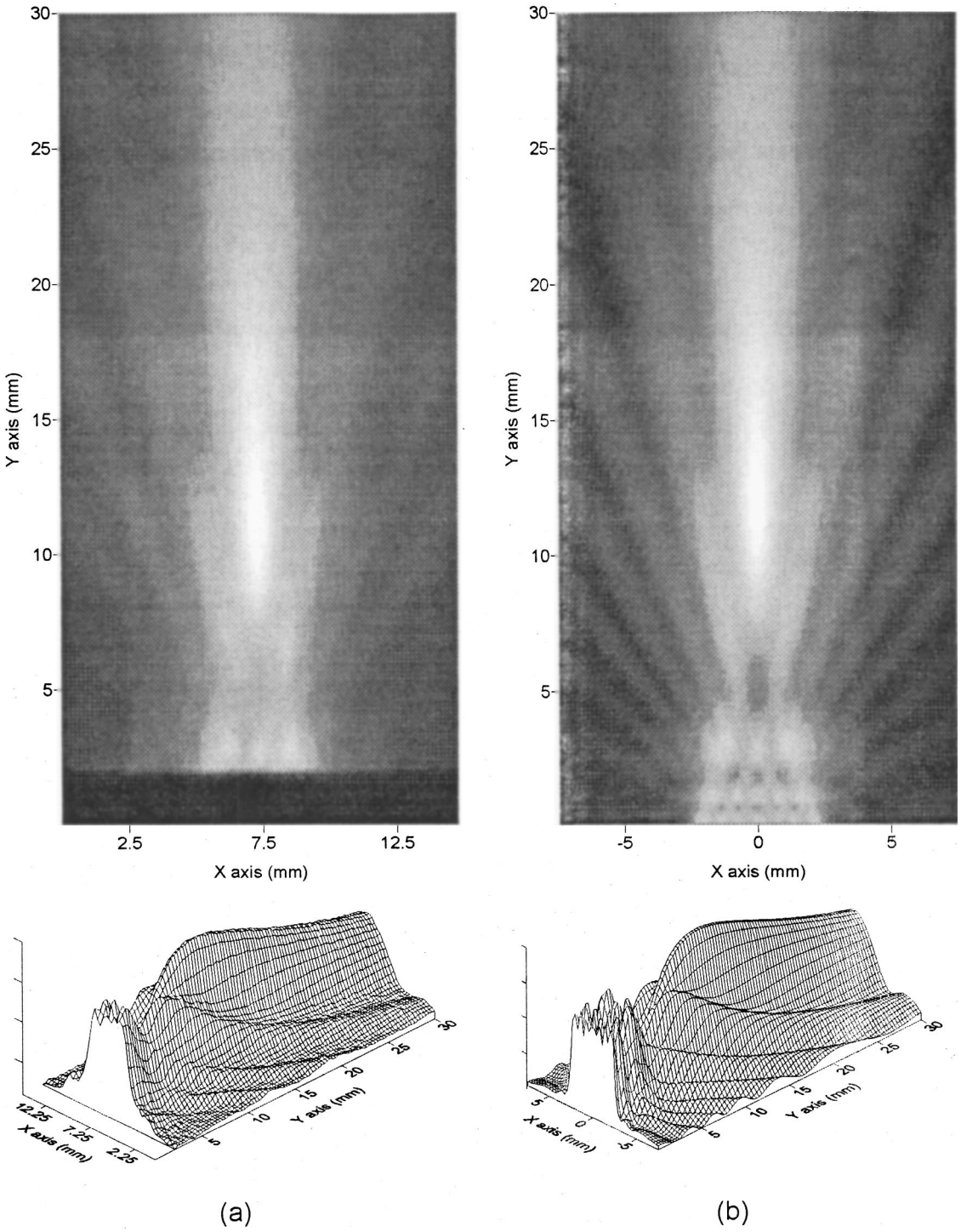


FIG. 12. Comparison of (a) experimental and (b) theoretical radiated fields for a 5-mm-square air–coupled cMUT driven with a 0.5-MHz tone-burst excitation signal.

follow the predicted fields very closely. This, in turn, suggests that the square cMUTs, when operated in air, were a very good approximation to plane piston radiators. This agrees with the conclusions reached by Safaeinili *et al.*[23] for polymer-film transducers used in the far field.

B. Tone-burst generation signals

The theoretical model can also be used to predict the acoustic field generated when other acoustic generation signals are used. Hence, such fields were investigated experimentally, using a 10-cycle tone-burst signal at 30 V from a Wavetek model 191 signal generator to excite the cMUT. This voltage level was found to be sufficient to detect signals in air with a good signal to noise ratio.

Figures 12–14 show the measured and predicted fields for a 5-mm-square cMUT, when driven by 0.5-, 1.0-, and 1.5-MHz tone-burst signals. Using the maximum output of the signal generator of 30 V, a bias of 8 V was found to be sufficient to prevent frequency doubling in experimental fields. The theoretical predictions again assumed square

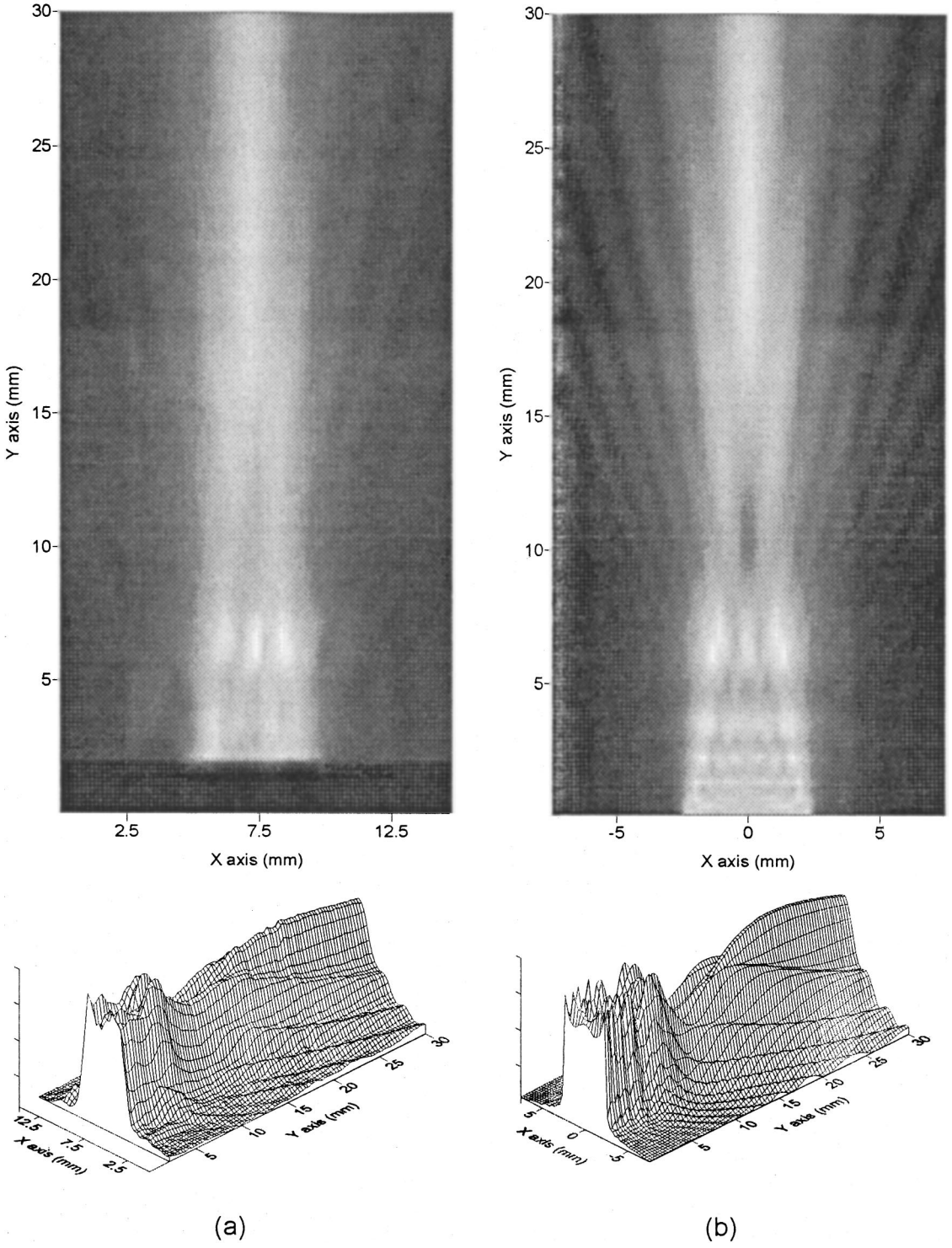


FIG. 13. Comparison of (a) experimental and (b) theoretical radiated fields for a 5-mm-square air–coupled cMUT driven with a 1.0-MHz tone-burst excitation signal.

plane pistons, driven with a 10-cycle tone-burst excitation signal with an amplitude of 30 V. It is evident in both theory and experiment that the radiated field is much more complicated, as expected, with a series of maxima and minima throughout the radiated field, and a number of sidelobes in the far field. As the frequency of the tone burst was increased, the sidelobes can be seen to converge, and the position of the near-field/far-field boundary can be seen to shift further from the face of the transducer, as expected. As in the case of impulsive excitation, the field plots of Figs. 12–14 confirm that the theoretical model gives close agreement to the experimental field characteristics emitted by these cMUTs, and again it can be concluded that the cMUTs are behaving in a piston-like manner.

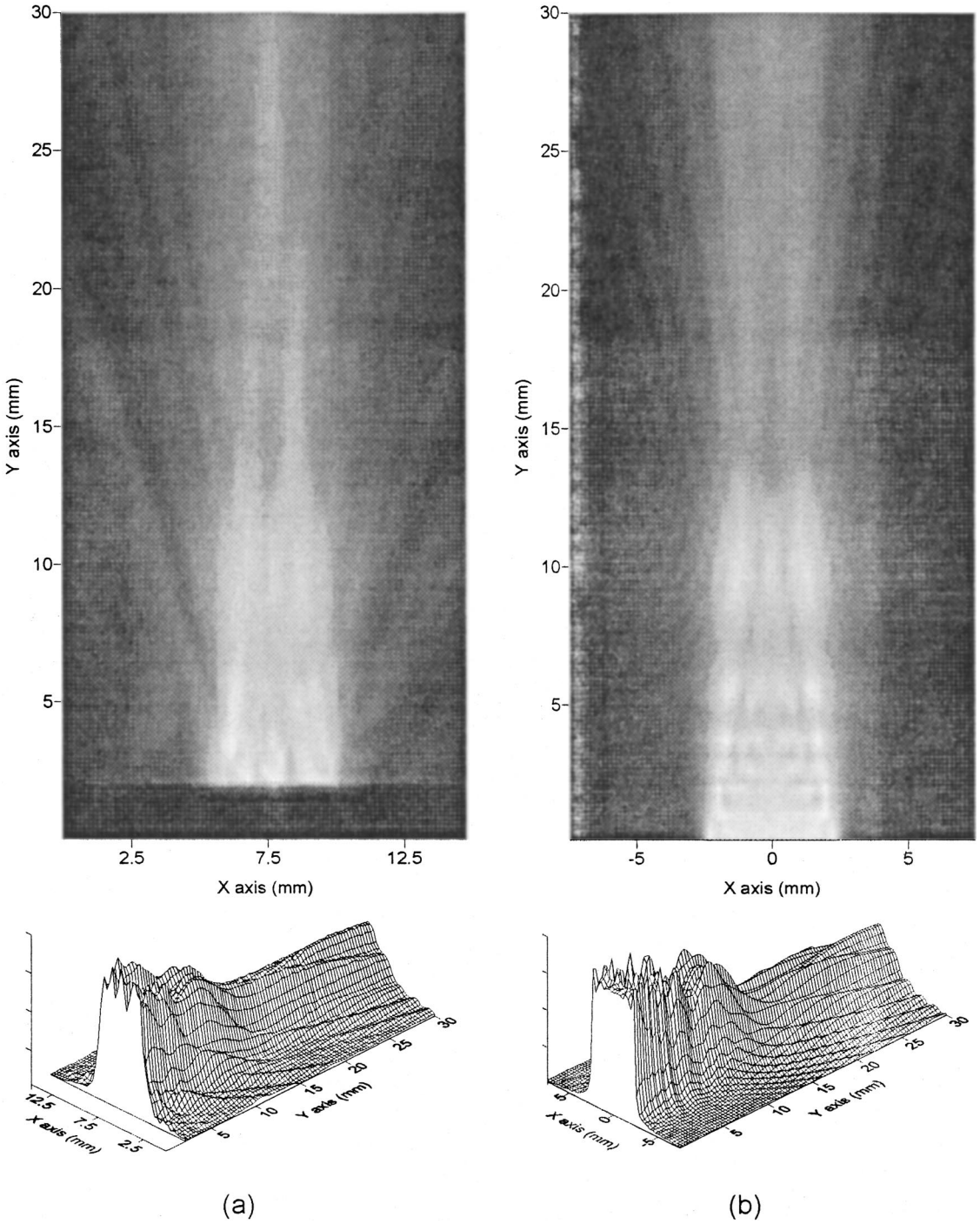


FIG. 14. Comparison of (a) experimental and (b) theoretical radiated fields for a 5-mm-square air–coupled cMUT driven with a 1.5-MHz tone-burst excitation signal.

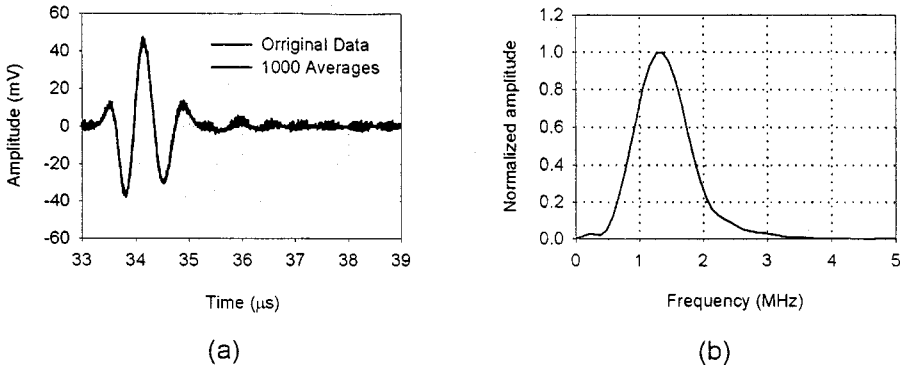


FIG. 15. (a) A typical waveform and (b) frequency spectrum, for a pair of cMUTs operating on axis in a pitch–catch mode in air.

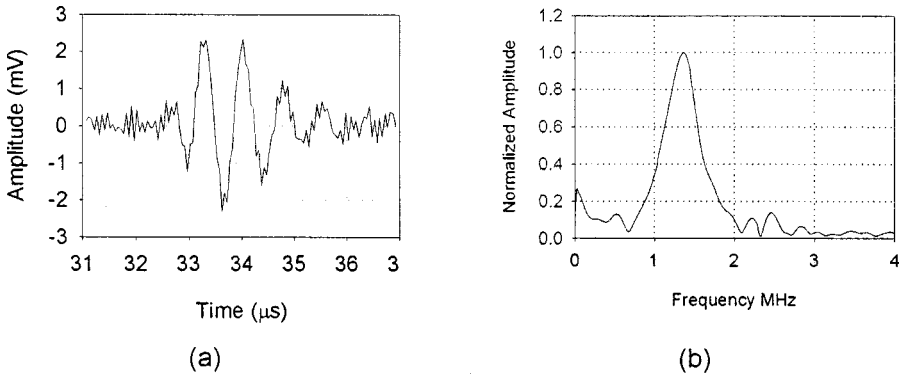


FIG. 16. (a) A typical waveform and (b) frequency spectrum for a cMUT operating in a pulse–echo mode.

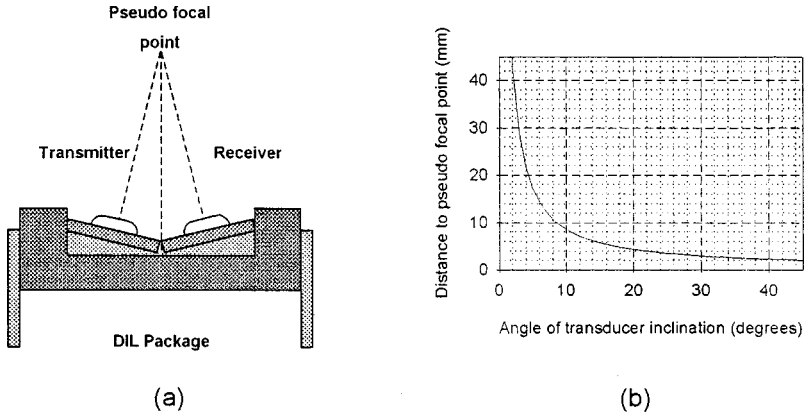


FIG. 17. (a) Schematic diagram of an imaging system using two devices mounted in a DIL package.

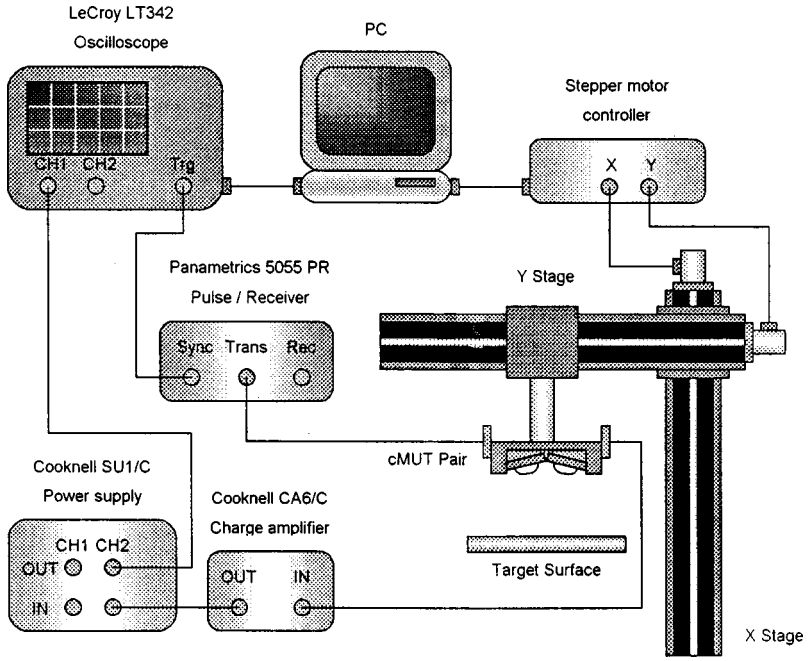


FIG. 18. Schematic diagram of a surface metrology system using pitch–catch cMUTs.

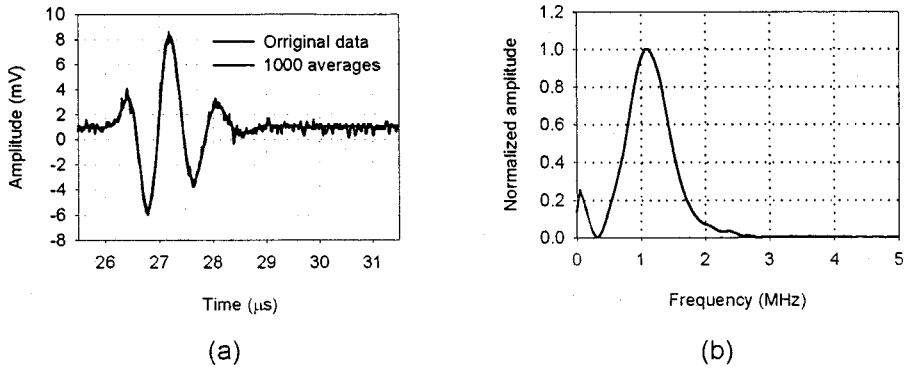


FIG. 19. Typical (a) waveform and (b) frequency spectrum for a pair of cMUTs operating in pulse–echo mode utilizing an impulse drive signal.

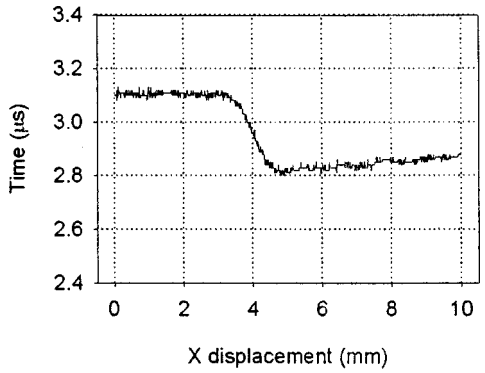


FIG. 20. The variation in ultrasonic arrival time as the transducer pair of Fig. 16 was scanned over a 50-μm step in a steel surface.

IV. USE OF A PAIR OF TRANSDUCERS FOR SURFACE MEASUREMENTS

It has been shown that the fields generated by these cMUTs are very predictable, and hence a series of experiments has been performed in which the devices were used for ultrasonic imaging and measurements of solid surfaces. Much work on the subject of ultrasonic imaging has been conducted, and many different strategies have been used. These strategies tend to fall into one of three primary categories, pulse–echo imaging (e.g., echo location), intensity mapping (e.g., medical imaging), and phase-amplitude measurement (e.g., volumetric imaging).[24] This imaging can use either discrete scanned transducers, or arrays. The former approach will be used here for imaging in air. With discrete transducers, images can be obtained by scanning either a source–receiver pair of transducers (the so-called pitch–catch mode), or with a single device acting in the pulse–echo mode.[25] A variation in one or more of the received signal properties (e.g., amplitude variation.) can be plotted against transducer position to create the image. This whole system falls into the category of reflected amplitude mapping.

Such a mapping system has been demonstrated using two 1-mm-square cMUTs, each with 1-μm-thick membranes, which were aligned axially at a 10-mm separation in air. As in the case of impulsive field scans, the transmitter was connected to the Panametrics 5055PR pulser–receiver, while the receiver output was fed to the Cooknell CA6/C charge amplifier/Cooknell SU1/C power supply combination. A representative waveform and corresponding frequency spectrum is shown in Fig. 15, which shows a reasonably damped signal with an amplitude of about ± 40 mV. The center frequency is 1.2 MHz, and the bandwidth is 1.0 MHz, giving a 6-dB bandwidth of around 80%. It can be seen in the results of Fig. 15 that the signal typically had 2 mV of high-frequency noise component, but this was easily removed by signal averaging, as shown.

In order to use a single device in this mode of operation, the transducer must be connected to both transmit and receive electronics simultaneously. The problem with this requirement is that the sensitive receiver system needs protection from the high-voltage excitation waveform from the Panametrics 5055PR. This requirement was realized with the use of a specially developed diplexer circuit. A single 1-mm-square cMUT was aligned axially 5 mm from a flat-mirrored surface (thus giving a total distance traveled of 10 mm), and used in pulse–echo mode with the diplexer. A representative waveform was recorded, and the frequency spectrum determined, as shown in Fig. 16. The two main differences between this signal and that of Fig. 15 were the amplitude of the signal and the signal-to-noise ratio. The damping and center frequency were both similar to those of Fig. 15. The bandwidth of the pulse–echo situation is a little lower; this was due to the lower amplitude of the received signal with respect to noise, caused mainly by the diplexer design which had a gain of ~ 0.1. A second reason for this lack of amplitude was that the applied dc voltage increases receiver sensitivity, but reduces that in ultrasonic generation.[26] Hence, a trade-off is required which reduces sensitivities further. However, operation as a pitch–catch pair yields no such problem because the two devices are independently biased.

A pitch–catch system has been developed for surface metrology measurements, using the arrangement shown schematically in Fig. 17(a). The radiated fields measured previously have shown that if two 1-mm-square cMUTs are mounted on separate substrates so that they can be angled towards each other, then the fields will overlap, resulting in a pseudofocal point. The transducers were thus angled towards each other as shown, and mounted in a standard electronics mounting system (a DIL package). A 200-V transient impulse signal was used as the generation signal for these measurements, with a 30-V bias applied to the receiver and no bias to the transmitter. The angle relative to their common axis was chosen so that there was enough power in the signal for reception, but that the angle was not so great that the mechanical structure of the package interfered with the sample being scanned. Simple trigonometry was used to generate the curve of Fig. 17(b). This shows the distance from

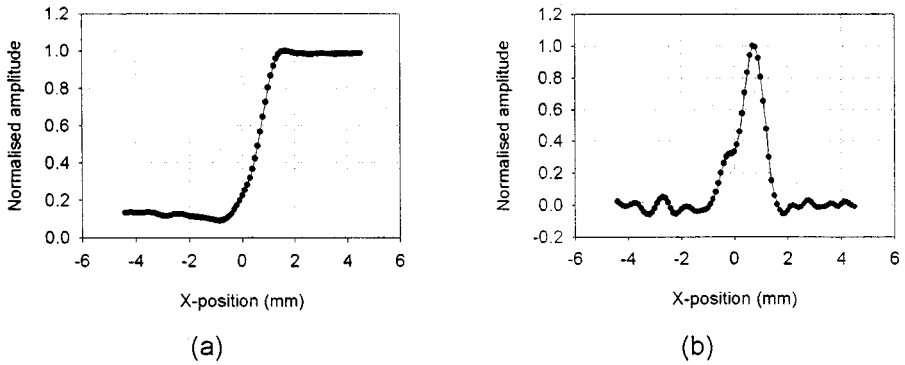


FIG. 21. (a) The measured horizontal resolution of a transmit–receive pair of cMUTs when fired with an impulse signal, when scanned over a knife edge, and (b) the first-order spatial derivative.

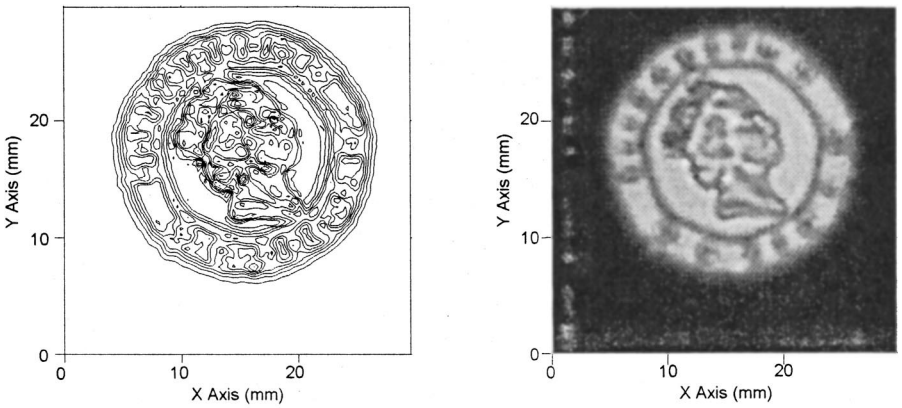


FIG. 22. (a) An image of a 20-pence British coin, generated from ultrasonic reflected amplitude data in air, using a pair of cMUTs in pitch–catch mode.

the center of the package to the pseudofocal point, plotted against the angle of inclination of the transducer. This assumes symmetry between the transducers. An angle of 10° was chosen, giving a region of interest some 8 mm away, while allowing the angle of incidence of the ultrasonic beam to be as close to zero as is practical. This helped to keep the spot size as small as possible, thus retaining maximum scan resolution.

The transducer pair was mounted on a linear stage, which could be scanned horizontally with a resolution of 0.01 mm, as shown in Fig. 18. The surface under test was aligned parallel to the scanning stage, and the distance to the sample was controlled by a second motorized stage. As an example, the signal recorded by reflection from a flat glass calibration block is shown in Fig. 19. It can be seen here that this signal is approximately 15 mV peak to peak, with a center frequency of about 1.2 MHz, but has about 1 mV peak to peak of high-frequency noise superimposed on the signal. As before, the simplest way of removing the majority of this noise from the signal is to use the averaging facility on the oscilloscope.

In order to determine the vertical resolution of this pulse–echo imaging system, a B scan was taken over slip gauges, which were arranged so as to form a series of steps of 20-, 50-, and 100-μm height difference. The signal arrival time was taken to be the first point crossing a negative threshold value. This was evaluated with custom signal-processing software. Figure 20 shows a typical result from such an experiment for a 50-μm step. It has been estimated from such measurements that the vertical height resolution is better than 20 μm. The horizontal resolution of this pulse–echo imaging system was also obtained [Fig. 21(a)] by scanning the transducer pair over a knife edge, using a scan resolution of 0.001 mm. This response was then differentiated, as shown in Fig. 21(b), and the FWHM (full width half maximum) resolution was found to be $\sim$1.5 mm. (Note: this is a standard test used in the determination of image resolution in imaging systems.) This is approximately 5λ in air at the center frequency of these devices ($\lambda = 283$ μm for a velocity of 340 ms^{-1} and a frequency of 1.2 MHz).

In order to examine the viability of a C-scan surface profile imaging system, a scan of the surface of a 20-pence British coin was performed. A scanning system was established so that an $X-Y$ horizontal raster scan could be performed across the surface of the coin. The scan resolution was 0.25 mm, and the scan dimensions were 30 by 30 mm. The data were again processed and an image was formed using the peak-to-peak amplitude of the received signals. Figure 22(a) shows the results of this scan, where the main features of the coin are visible. The resolution of this system has room for improvement in several areas. The accuracy of the mounting of the cMUTs can be improved, perhaps by machining a pair of wedges on which to mount the cMUTs. It may also be possible to use an array of small elements to steer the beam to the focal point, focusing the beam to a more highly convergent spot, thus increasing the horizontal resolution, and possibly the vertical resolution (within the constraint of the length of the focal spot). The inherent noise in the system could be reduced, and the system could be improved in terms of speed by changes in hardware and software. These improvements are currently being investigated.

V. CONCLUSIONS

It has been demonstrated that theory for the field of a rectangular plane piston can be used to predict the actual field of cMUTs radiating into air. For this to occur, the theoretical approach needs to account for ultrasonic attenuation in air, for example by using the model proposed by Bond.[21] It has also been demonstrated that miniature detectors, also fabricated using the same technology, can be used for effective field scanning. In this paper, a 500-μm receiver was used, although future research will investigate even smaller devices. However, when combined with a suitable translational stage, such a receiver could be used to map the acoustic field of many transducers and acoustic configurations.

With knowledge of the radiated field, it was also possible to use a pair of cMUTs operating in a pitch–catch mode to image surface features. This approach has possible application to surface metrology, where contact to the surface being measured is not desirable.

[1] M. I. Haller and B. T. Khuri-Yakub, ''A surface micromachined electrostatic ultrasonic air transducer,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 1–6 (1996).
[2] I. Lådabaum, B. T. Khuri-Yakub, D. Spoliansky, and M. I. Haller, ''Mi-

cromachined ultrasonic transducers (MUTs),'' in Proceedings of the 1996 IEEE Ultrasonics Symposium, pp. 535–538 (1996).
[3] I. Ladabaum, B. T. Khuri-Yakub, and D. Spoliansky, ''Micromachined ultrasonic transducers: 11.4 MHz transmission in air and more,'' Appl. Phys. Lett. **68**, 7–9 (1996).
[4] P. C. Eccardt, K. Niederer, T. Scheiter, and C. Hierold, ''Surface micromachined ultrasound transducers in CMOS technology,'' in Proceedings of the 1996 IEEE Ultrasonics Symposium, pp. 959–962 (1996).
[5] P. C. Eccardt, K. Nieder, and B. Fischer, ''Micromachined transducers for ultrasound applications,'' in Proceedings of the 1997 IEEE Ultrasonics Symposium, pp. 1609–1618 (1997).
[6] P. C. Eccardt and K. Niederer, ''Micromachined ultrasound transducers with improved coupling factors from a CMOS compatible process,'' Ultrasonics **38**, 774–780 (2000).
[7] G. Caliano, V. Foglietti, E. Cianci, and M. Pappalardo, ''A silicon microfabricated electrostatc transducer: 1 MHz transmission in air and water,'' Microelectron. Eng. **53**, 573–576 (2000).
[8] W. P. Mason, *Electromechanical Transducers and Wave Filters* (Van Nostrand, New York, 1942).
[9] R. A. Noble, R. J. Bozeat, T. J. Robertson, and D. A. Hutchins, ''Novel silicon nitride micromachined wide bandwidth ultrasonic transducers,'' in Proceedings of the 1998 IEEE Ultrasonics Symposium, pp. 1081–1084 (1998).
[10] T. J. Robertson, A. Neild, D. A. Hutchins, J. S. McIntosh, D. R. Billson, R. A. Noble, R. R. Davies, and L. Koker, ''Radiated fields of rectangular air-coupled micromachined transducers,'' in Proceedings of the 2001 International IEEE Ultrasonics Symposium, pp. 891–894 (2001).
[11] G. R. Harris, ''Review of transient field theory for a baffled planar piston,'' J. Acoust. Soc. Am. **70**, 10–21 (1981).
[12] J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. 1.
[13] P. R. Stepanishen, ''Transient radiation from pistons in an infinite planar baffle,'' J. Acoust. Soc. Am. **50**, 1629–1638 (1971).
[14] J. C. Lockwood and J. G. Willette, ''High speed method for computing the exact solution for the pressure variations in the nearfield of a baffled piston,'' J. Acoust. Soc. Am. **53**, 735–741 (1972).
[15] J. L. San Emeterio and L. G. Ullate, ''Diffraction impulse response for rectangular transducers,'' J. Acoust. Soc. Am. **92**, 651–662 (1992).
[16] A. Freedman, ''Sound field of a rectangular piston,'' J. Acoust. Soc. Am. **32**, 197–209 (1960).
[17] A. Neild, ''Advances in Air Coupled Ultrasonic Arrays,'' Ph.D. thesis in Engineering, University Of Warwick, U.K. (2003).
[18] T. J. Robertson, ''Advances in Ultrasonic Capacitive Transducer Technology,'' Ph.D. thesis in Engineering, University Of Warwick, U.K. (2001).
[19] L. G. Ullate and J. L. S. Emeterio, ''Diffraction impulse response of rectangular transducers,'' J. Acoust. Soc. Am. **92**, 651–662 (1992).
[20] L. G. Ullate and J. L. S. Emeterio, ''A new algorithm to calculate the transient near-field of ultrasonic phased arrays,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 745–753 (1992).
[21] L. J. Bond, C.-H. Chiang, and C. M. Fortunko, ''Absorption of ultrasonic waves in air at high frequencies,'' J. Acoust. Soc. Am. **92**, 2006–2015 (1992).
[22] O. I. Lobkis, A. Safaeinili, and D. E. Chimenti, ''Precision ultrasonic reflection studies in fluid-coupled plates,'' J. Acoust. Soc. Am. **99**, 2727–2736 (1996).
[23] A. Safaeinili and D. E. Chimenti, ''Air-coupled ultrasonic estimation of viscoelastic stiffness in plates,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 1171–1180 (1996).
[24] G. Wade, ''Human uses of ultrasound: Ancient and modern,'' Ultrasonics **38**, 1–5 (2000).
[25] P. S. Naidu, *Sensor Array Signal Processing* (CRC Press, Boca Raton, 2001).
[26] J. S. McIntosh, D. A. Hutchins, D. R. Billson, T. J. Robertson, and R. A. Noble, ''The characterization of capacitive micromachined ultrasonic transducers in air,'' Ultrasonics **40**, 477–483 (2002).

Measurements of the longitudinal wave speed in thin materials using a wideband PVDF transducer

Kwang Yul Kim, Wei Zou, Steve Holland, and Wolfgang Sachse
Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853

(Received 7 September 2001; accepted for publication 30 January 2003)

A flat transducer was constructed, using a 9-μm-thick PVDF (polyvinylidene fluoride) film for generation and detection of high-frequency ultrasonic waves, and used for measurements of the phase velocity of longitudinal waves traveling along the thickness direction in a very thin material. The transducer has a useful wideband frequency characteristic extending from 10 MHz to over 150 MHz. Measurements of the phase velocity of the longitudinal waves are carried out using a 0.212-mm-thick glass slide and a 0.102-mm-thick stainless-steel shim, using water as a coupling medium. The thickness limit for this measurement appears to be approximately 20 μm. The phase velocity of the longitudinal mode is obtained as a function of frequency in the frequency domain by using a modified sampled continuous wave (cw) technique. It can also be measured in the time domain by using a broadband pulse of short duration. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1562650]

PACS numbers: 43.38.Ar, 43.38.Fx [SLE]

I. INTRODUCTION

Accurate ultrasonic measurements of sound speed along the thickness direction in a plate specimen are usually based on the interference phenomenon between the successive echoes as the ultrasonic beams reverberate inside the plate. The phase comparison method (Williams and Lamb, 1958), pulse superposition method (McSkimin, 1961; McSkimin and Andreatch, 1962), and continuous wave (cw) resonance method (Bolef and Miller, 1971) are the typical examples that make use of the interference or alignment of the successive echoes, as one varies either the frequency of a tone-burst signal (pulse repetition rate) or its carrier frequency. A variation of these methods is the pulse–echo overlap method (Papadakis, 1976) in which best alignment or overlap of the successive echoes is sought to find their carrier frequency. Constructive and destructive interferences may occur at frequencies corresponding to the integral multiples of one half-wavelength. For a thin-plate specimen this requires a high-frequency transducer that generates and detects ultrasonic waves.

The spherically focusing (Smolorz and Grill, 1996) and line-focusing (Zou *et al.*, 2003) high-frequency transducers, which are both made of a 9-μm-thick piezoelectric PVDF (polyvinylidene fluoride) film, are reported to possess a wide-frequency bandwidth ranging from 10 MHz to over 150 MHz. Very recently, the authors (Zou *et al.*, 2003) demonstrated that a line-focusing PVDF transducer can be used to measure the phase velocity of a surface acoustic wave (SAW) in a $V(z)$ curve by employing either a sharp broadband pulse or a tone-burst signal having a carrier frequency chosen over a wide frequency region. It is quite natural to expect that a flat transducer made of the same piezoelectric PVDF film can be used to measure the phase velocity of longitudinal waves. Because of its high-frequency bandwidth, it can be applied to investigate very thin materials. We fabricated the flat PVDF transducer and used it both as a generator and as a detector of ultrasonic waves to successfully measure the phase velocity of the longitudinal mode propagating in the thickness direction of very thin glass slide and stainless-steel shim. Although both a broadband pulse and a tone burst of various carrier frequencies are used for this purpose, of particular interest is a modified sampled cw technique adapted to the flat PVDF transducer aligned parallel to a very thin-plate specimen using water as a coupling medium, as one sweeps frequency over the effective bandwidth of the PVDF transducer.

II. DESCRIPTION OF MODIFIED SAMPLED cw ULTRASONIC SPECTROSCOPY

A flat transducer using a 9-μm piezoelectric PVDF film is fabricated, following the design and construction technique similar to those adopted in a line-focusing transducer (Zou *et al.*, 2003), in which the geometry of a line-focusing part is simply replaced by the flat geometry. One flat transducer is used as a noncontact sensor to generate and detect ultrasonic waves, using water as a coupling medium to a thin specimen. We modified a sampled cw technique (Bolef and Miller, 1971) to observe a series of resonance frequencies corresponding to an integral multiple of half-wavelengths of ultrasound in the specimen as the frequency is swept over a wide bandwidth of the transducer.

An overall electronic block diagram for the measurement system is displayed in Fig. 1. For a radio frequency (rf) tone-burst operation, we used an electronic signal generator capable of generating both continuous and gated harmonic signals in the frequency range from 5 kHz to 1.5 GHz. The rf tone-burst signals derived from the signal generator are fed into a rf power amplifier, whose frequency bandwidth extends from 100 kHz to 250 MHz. The output of the rf amplifier is connected to one arm of the single-pole double-throw (SPDT) switch, in which the single pole always joins the PVDF transducer. The operation of the SPDT switch is controlled by the TTL pulse generator that is synchronized with the rf signal generator at the beginning of signal generation. The switching times of the SPDT switch are less

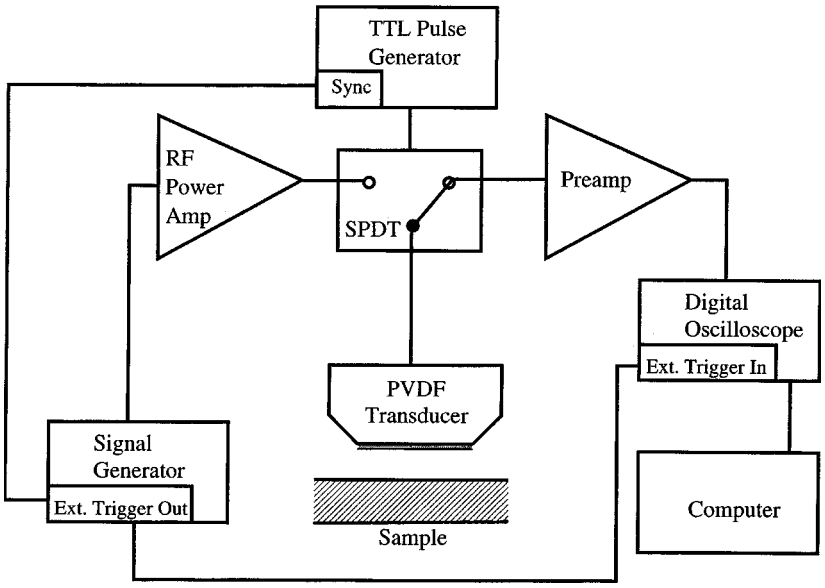


FIG. 1. Electronic block diagram for generation and detection of ultrasonic signals for the modified sampled cw technique.

than 20 ns and the width of the TTL pulse is maintained slightly longer than that of the rf tone-burst signals. During the time when the TTL pulses are on, the single pole of the SPDT switch is connected to the rf amplifier side arm of the switch, and the PVDF transducer generates longitudinal ultrasonic signals that transmit through water to a thin specimen. The ultrasonic signals are then reflected from the specimen surface, reverberate inside the specimen, and are transmitted back to the PVDF transducer.

In the modified sampled cw technique, a continuously running oscillator is gated on for a sufficiently long time, so that a steady-state cw acoustic response is established in the specimen. But, the gate pulse width is kept shorter than the round-trip time through water between the transducer and the specimen, so that signals generated and detected by the same transducer are separated from each other and not superposed together. After the steady-state conditions have been reached, the signal transmission gate is off and the signal detection gate is on. The distance between the PVDF transducer and the specimen is taken to be about 2 mm, which corresponds to an ultrasonic round-trip time of about 2.67 μs. We choose the gate width for the tone-burst signals to be 2.4 μs, which corresponds to slightly longer than 32 round-trip times of the longitudinal ultrasonic waves that undergo multiple reflections inside a 0.212-mm glass slide used as a thin specimen. The longer the gate width for tone-burst signals is, the sharper will be the mechanical resonance that is prompted by more reverberations in the specimen. The longer gating time, which is maintained slightly shorter than an ultrasonic round-trip time between the transducer and the specimen, requires a longer distance in water. However, attenuation in water increases in proportion to the square of frequency. As a result, high-frequency signals will attenuate severely as they travel a longer distance. The 2-mm water gap maintained between the transducer and the specimen appears to be an optimum distance, which allows a steady state to be set up in the specimen and still preserves high-frequency signals as they make a round trip through water from the transducer to the specimen. The pulse repetition rate chosen is 10 ms, which is long enough for the generated and reverberated signals to decay out almost completely before the next pulse for signal excitation is on.

During the time the TTL pulse is off, reflected and reverberated signals from the specimen arrive at the PVDF transducer, which via the single pole of the SPDT switch is in contact with the other arm of the switch connected to the input of a preamplifier whose bandwidth extends from dc to over 300 MHz. The SPDT switch prevents the preamplifier from being overloaded from the high-amplitude output of the rf amplifier, and keeps the electromagnetic leakage of a generated signal to a detected signal below 100 dB. The output of the preamplifier is brought into a digital sampling oscilloscope, which digitizes the signals at the sampling rate of 2.5 GHz and displays them on a phosphor screen for visual observation. Finally, the digitized signals are brought into a high-speed personal computer (PC) for processing and storage of the signals.

As the carrier frequency of tone-burst signals is swept, the shape of observed signals on the scope changes and displays a series of mechanical resonances evenly spaced in the frequency domain for nondispersive materials. Figure 2 shows two signals observed with a 0.212-mm glass slide, one at an off-resonance frequency of 127 MHz and the other at a resonance frequency of 120.6 MHz. For most of the frequency range corresponding to the off-resonance condition, signals similar to the upper one in Fig. 2 are observed. As the swept frequency approaches towards a resonance frequency in the narrow interval surrounding the resonance frequency, the trailing edge of the signal rapidly builds up in amplitude, which decays in time. The peak amplitude among the trailing edge signals occurs at a resonance frequency. To enhance the difference between the off-resonance and near-resonance signals, the magnitude of a trailing signal for 0.5-μs duration immediately after the termination of the excitation signal is integrated. These integrated values are plotted as a function of frequency and shown in Fig. 3, which exhibits sharp resonance peaks at an evenly spaced frequency interval (13.40 MHz) in the frequency range from 10 to 150 MHz. The sharp

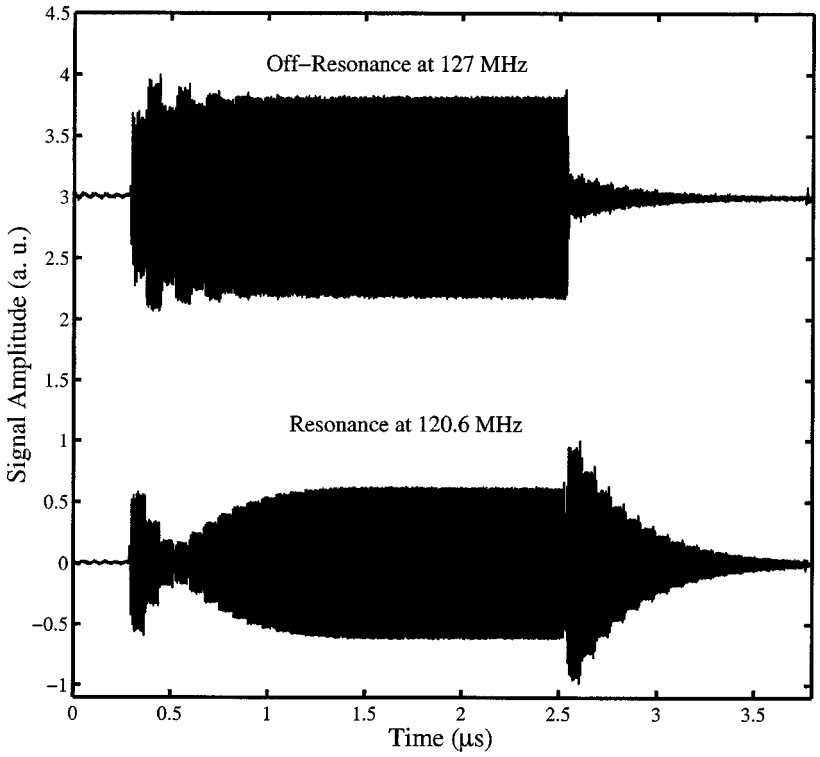


FIG. 2. Signals observed with a 0.212-mm-thick glass slide at off-resonance and resonance frequencies.

resonance peaks at the center of a mechanical resonance occur at frequencies corresponding to the integral multiples of one half-wavelength of ultrasounds as multiple reflections in the specimen constructively superpose to set up standing waves. The signal shapes shown in Fig. 2 and the overall shape of the envelope of the frequency domain sweep in Fig. 3 are discussed in detail in the literature (Bolef and Miller, 1971). From the evenly spaced frequency interval 13.4 MHz between successive resonance peaks in the 0.212-mm-thick glass slide specimen, the longitudinal phase velocity of 5.68 km/s is obtained for glass. We conducted a similar experiment using a 0.102-mm-thick stainless-steel shim, which displays resonance peaks evenly spaced at a frequency interval of 28.75 MHz. From this value, we obtain the phase velocity 5.84 km/s of the longitudinal mode for stainless steel. Supposing that the first resonance peak can be observed near 140 MHz, the thickness limit of the specimen materials for measurement of the longitudinal phase velocity appears to be about 20 μm.

The frequency sweep is implemented with an interval of 0.1 MHz between 10 and 150 MHz at a total of 1401 frequencies. At each frequency 10 000 digitized signal points are brought to a computer at a sampling rate of 2.5 GHz. The resonance frequencies corresponding to the peaks in Fig. 3

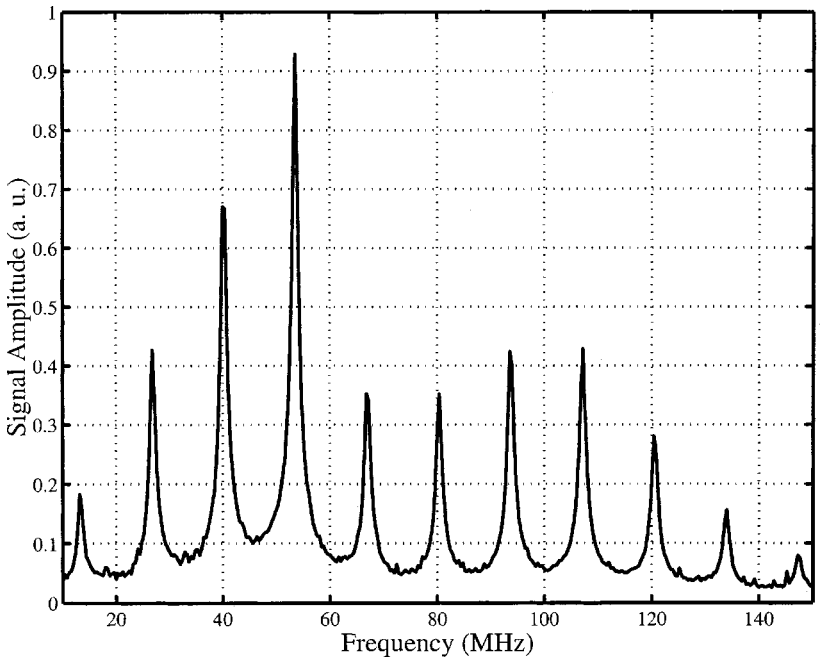


FIG. 3. Resonance spectroscopy of a 0.212-mm-thick glass slide in the frequency domain.

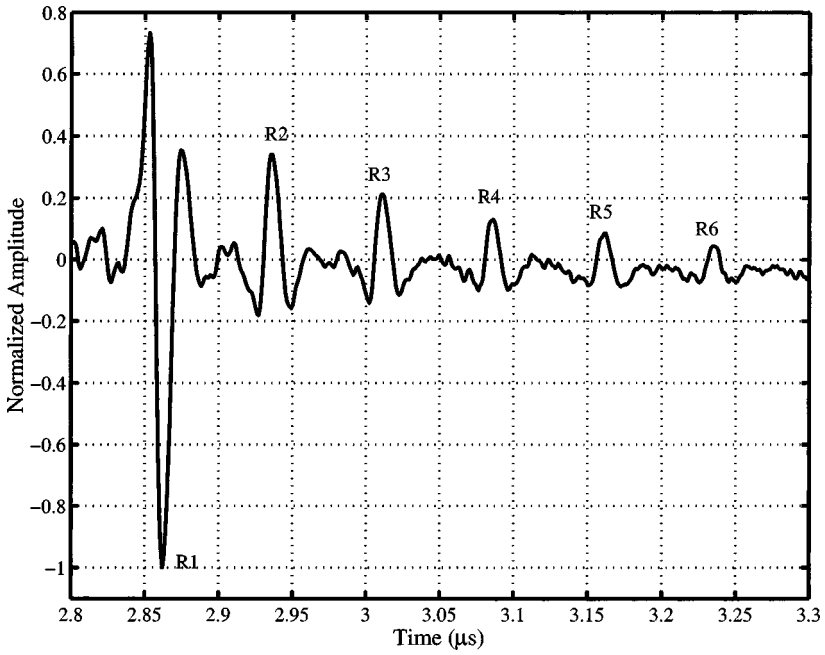


FIG. 4. Echoes observed from a 0.212-mm-thick glass slide, using a broadband sharp pulse.

are thus determined with an error $\leqslant 0.05$ MHz over the frequency span over 100 MHz. As a result, an error in transit time measurement of the longitudinal wave across the specimen is better than 0.05%. The error in transit time measurements can be improved by taking a smaller interval in frequency sweep, say 0.01 MHz, at the cost of bringing more digitized points in computer storage. However, the thickness of a specimen is determined with an error of 1 μm, and therefore the errors in measurements of the longitudinal phase velocity both in glass and in stainless-steel specimens are estimated to be about 0.5% and 1%, respectively.

III. BROADBAND PULSE RESPONSE

For nondispersive materials, such as glass and stainless steel used as a specimen material, a broadband pulse generated and detected by the flat PVDF transducer can be used to measure in the time domain the phase velocity of longitudinal waves traveling a thin-plate specimen. For this purpose, we used the 200-MHz-bandwidth pulser/receiver (Panametrics model 5900PR). The broadband pulse response of the system, obtained by using 5-GHz sampling rate and the 0.212-mm-thick glass slide, is shown in Fig. 4, which displays multiple reflections undergone in the glass specimen. In the figure, R1 is the first echo reflected from the specimen surface facing the transducer, and R2, R3, R4, etc., are the echoes which are first transmitted through the specimen boundary facing the transducer, which make then one round trip, two round trips, three round trips, etc., respectively, in the specimen, and which are finally transmitted back to the transducer. Note that the phases of R1 and the rest of the echoes are reversed. The rest of the echoes are observed at a regular time interval, from which we obtain the phase velocity 5.68 mm/μs of the longitudinal mode in glass. This value is identical to the phase velocity measured by the modified sampled cw technique.

IV. CONCLUSIONS

We have demonstrated that a broadband, high-frequency, flat PVDF transducer, which is constructed with a 9-μm-thick piezoelectric PVDF film, can be used to measure the phase velocity of the longitudinal waves traveling along the thickness direction in a very thin plate. A modified sampled cw technique using gated tone-burst harmonic signals, especially adapted to the noncontact configuration of the transducer, works successfully to show sharp resonance peaks in the frequency-domain spectroscopy. For nondispersive materials, a broadband, sharp pulse can also be used for the flat PVDF transducer to measure the longitudinal phase velocity in the time domain.

ACKNOWLEDGMENTS

The financial support of the Physical Acoustics Program of Office of Naval Research is greatly appreciated. The use of the facilities of the Cornell Center for Materials Research (CCMR) of Cornell University is also acknowledged.

Bolef, D. I., and Miller, J. G. (**1971**). ''High Frequency Continuous Wave Ultrasonics'', in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York), Vol. 8, pp. 95–201.

McSkimin, H. J. (**1961**). ''Pulse Superposition Method for Measuring Ultrasonic Wave Velocities in Solids'', J. Acoust. Soc. Am. **33**, pp. 12–16.

McSkimin, H. J., and Andreatch, P. (**1962**). ''Analysis of the Pulse Superposition Method for Measuring Ultrasonic Wave Velocities as a Function of Temperature and Pressure'', J. Acoust. Soc. Am. **34**, pp. 609–615.

Papadakis, E. P. (**1976**). ''Ultrasonic Velocity and Attenuation: Measurement Methods with Scientific and Industrial Applications'', in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York), Vol. 12, pp. 277–374.

Smolorz, S., and Grill, W. (**1996**), ''Focusing PVDF transducers for acoustic microscopy'', Res. Nondestruct. Eval. **7**, pp. 195–201.

Williams, J., and Lamb, J. (**1958**). J. Acoust. Soc. Am. **30**, 308.

Zou, W., Holland, S., Kim, K. Y., and Sachse, W. (**2003**). ''Wideband high-frequency line-focus PVDF transducer for materials characterization'', Ultrasonics **41**, pp. 157–161.

Optimization of structural variables of a flextensional transducer by the statistical multiple regression analysis method

Kookjin Kang
Department of Sensor Engineering, Kyungpook National University, 1370 Sankyukdong, Bukgu, Daegu 702-701, Korea

Yongrae Roh[a)]
School of Mechanical Engineering, Kyungpook National University, 1370 Sankyukdong, Bukgu, Daegu 702-701, Korea

(Received 22 November 2002; revised 8 May 2003; accepted 23 June 2003)

The performance of an acoustic transducer is determined by the effects of many structural variables, and in most cases the influences of these variables are not linearly independent of each other. To achieve optimal performance of an acoustic transducer, we must consider the cross-coupled effects of its structural variables. In this study, with the finite-element method, the variation of the operation frequency and sound pressure of a flextensional transducer in relation to its structural variables is analyzed. Through statistical multiple regression analysis of the results, functional forms of the operation frequency and sound pressure of the transducer in terms of the structural variables were derived, with which the optimal structure of the transducer was determined by means of a constrained optimization technique, the sequential quadratic programming method of Phenichny and Danilin. The proposed method can reflect all the cross-coupled effects of multiple structural variables, and can be extended to the design of general acoustic transducers. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1600725]

PACS numbers: 43.38.Fx, 43.30.Yj, 43.58.Fm [AJZ]

I. INTRODUCTION

A class IV flextensional transducer is a typical underwater acoustic transducer capable of generating high-power sound waves. Much work has been done on its operation principle and structure.[1–3] Piezoceramic stacks inside the transducer induce a small extensional displacement in the major axis direction of the transducer when exposed to electrical drive. The small displacement along the major axis causes a large displacement along the minor axis by the leverage effect of a compliant shell, and this provides the large volume displacement for high-power acoustic waves. For this flextensional transducer, the two most important performance factors are its operation frequency and the transmitted sound-pressure level.[4] Operation frequency in this paper is defined as the frequency at which a transducer shows the maximum sound pressure. The flextensional transducer is normally considered as an omnidirectional radiator because its size is small compared with the acoustic wavelength at its operation frequency range.[2,5]

The achievement of a high sound-pressure level has been made possible either by the use of various active materials such as electrostrictive and magnetostrictive materials or by the optimization of the transducer structure. A great deal of research has been carried out to discover the optimal structural configuration of the flextensional transducer.[6–10] However, most work has been done by considering the effects of specific variables of the transducer. The performance of an acoustic transducer is influenced by many structural variables,[9,10] and in general the influences of these variables are not linearly independent of each other.[9,12] A design method to consider all the cross-coupled effects of structural variables is rare because investigation of the coupled effects using a mathematical analytic method is quite difficult. Furthermore, investigation of all the effects through finite-element analyses (FEA) is very time consuming, and is almost prohibitive for a transducer with many structural variables like the flextensional transducer considered in this paper.

In this study, we have developed a new design scheme that can determine the optimal geometry of a class IV flextensional transducer capable of transmitting acoustic waves of the highest possible pressure level at a given operation frequency. The design scheme can reflect not only the individual, but also all the cross-coupled effects of the structural variables. In comparison with conventional analytic and finite-element methods, our method can determine the detailed geometry of a transducer with great efficiency and rapidity. Moreover, the technique is so general that it can be applied to any other type of acoustic transducers as well.

Initially, with the finite-element method (FEM), we analyzed the variation of the operation frequency and the sound pressure of the transducer in relation to its main structural variables. The structural variables to be analyzed were selected according to the results of a literature survey and our preliminary calculation of their effects on the transducer performance.[9,11] The transducer was composed of the materials in typical use for a flextensional transducer,[11] and the effects of material properties were not considered here. Through statistical multiple regression analysis of the FEA results on the performance variation, we formulated both the operation frequency and the sound pressure of the transducer

[a)]Electronic mail: yryong@knu.ac.kr

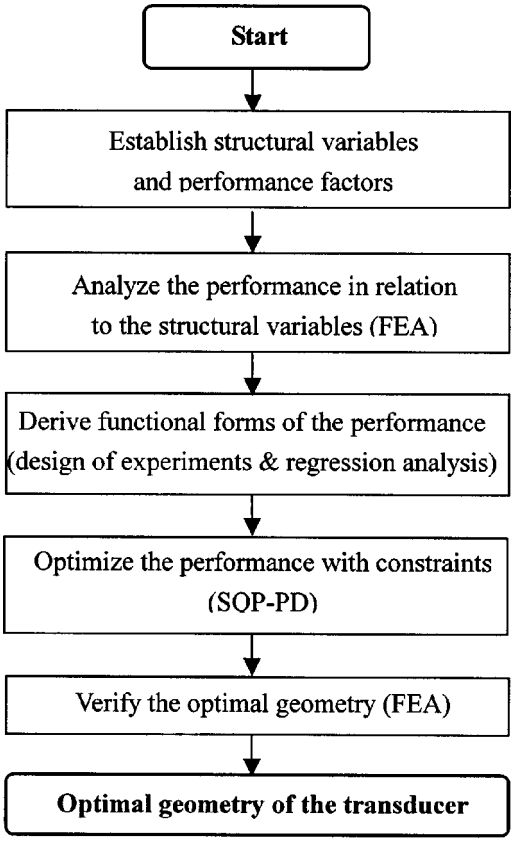


FIG. 1. Optimal design algorithm for the transducer.

as a function of the structural variables.[13,14] With the functional form of the sound pressure constrained by the desired operation frequency range and the dimension of the transducer, we determined the optimal combination of the structural variables to achieve the highest sound-pressure level by means of the constrained optimization technique, sequential quadratic programming method of Phenichny and Danilin (SQP-PD).[15] The whole design scheme is summarized in Fig. 1. The validity of the scheme was illustrated with the sample design of a flextensional transducer working at 1 kHz.

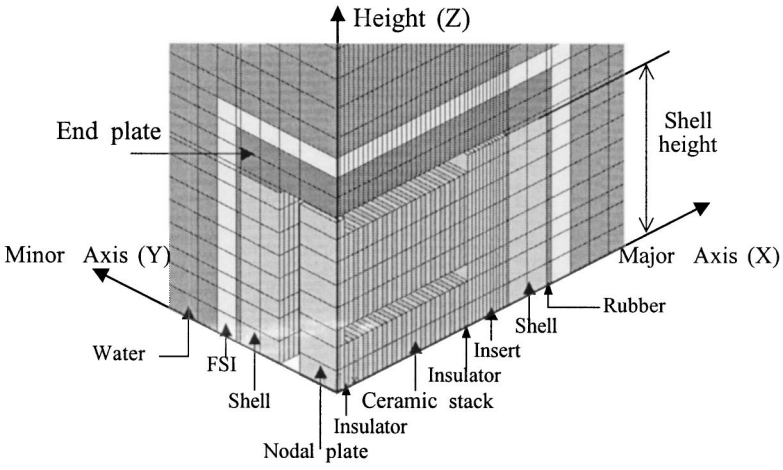
II. FINITE ELEMENT MODEL

As shown in Fig. 2, a three-dimensional finite-element model of a class IV flextensional transducer was constructed, which included all the details of the transducer. Considering symmetry conditions, only one-eighth of the whole transducer was modeled. The modeling and analysis were carried out with a commercial FEA package, ANSYS. To prevent reflection of an acoustic wave by the fluid medium, infinite boundary (pressure relief) conditions were applied to all the outside fluid boundaries as shown in Fig. 3. The transducer comprises an elliptical shell that acts as an acoustic radiator, and piezoelectric ceramic stacks that in turn act as active drivers. A compliant rubber layer covers the shell to work as an acoustic impedance matching layer as well as a protection layer. The nodal plate is at the center of the transducer, and the insulators are at each end of the piezoceramic stacks to electrically isolate the shell from the ceramic stacks. The ceramic stacks are connected to the shell by means of metallic inserts. A pair of metallic end plates seals the open ends of the shell. The width of the ceramic stacks is large enough to contact the interior surface of the shell as described in Fig. 2(b). Hence, the stack width changes in proportion to the ratio of the minor axis length to the major axis length of the shell. The height of the ceramic stacks was set to occupy 80% of the shell height, which is the common ratio in use.[1,16] The constructed FEM model including the acoustic radiation medium, water, consists of about 25 000 nodes and 22 000 elements. In order to verify the validity of the FE model, we analyzed vibration modes of the transducer. The result, illustrated in Fig. 4, is a clear first flextensional vibration mode of the shell that agrees well with previously reported results.[17,18] In general, the operation of a flextensional transducer makes use of the combinatorial motion of the first flextensional mode of the shell and the first length extensional mode of the ceramic stack.[17]

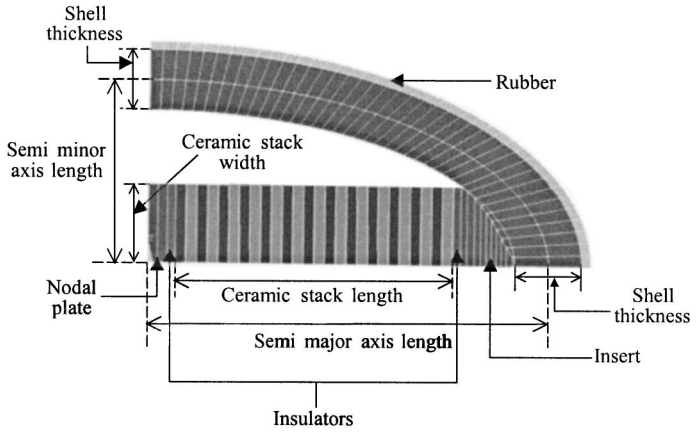
In this study, the transducer in Fig. 2 was designed to work at 1 kHz. Through preliminary calculations, the basic geometry of the transducer exhibiting an operation frequency of 1 kHz was determined as shown in Table I. Table II shows the properties of the materials composing the transducer, which were determined following the common structure of a flextensional transducer.[1,2,4] Once the FE model was constructed, the process to determine the optimal geometry of the transducer was conducted following the routine in Fig. 1. Each of the following sections describes each step of the optimization process. First, following the results of the preliminary calculations on the performance of the transducer in relation to its structural parameters shown in Fig. 2, the most influential structural variables were established: the semimajor axis length, the ratio of the minor axis length to the major axis length, the height and the thickness of the shell, and the length of the ceramic stacks. Performance factors are the radiated sound pressure and the operation frequency of the transducer. The sound pressure works as a target to be maximized and the desired operation frequency works as a constraint on the design of the transducer.

III. ANALYSES OF THE OPERATION FREQUENCY AND THE SOUND PRESSURE OF THE TRANSDUCER

Through FE harmonic analyses, we investigated the change of the operation frequency and the sound pressure of the transducer in relation to the variation of its structural variables. The variation ranges for the structural variables, as shown in Table III, had wide enough spans to include all the possible modifications of the variables for practical applications. We started with a model having the basic geometry in Table I. Then, structural variables were varied one by one over the range in Table III, while all other variables were fixed to the basic geometry. As an illustrative result, Fig. 5 shows typical harmonic responses from the FEA, and shows the variation of the transducer performance in relation to the change of semimajor axis length. We can determine the maximum sound pressure for each length of the semimajor axis, and the operation frequency at which the sound pressure shows the maximum value. Electrical voltage applied to the piezoceramic stack was arbitrarily set to be 1 V/mm all through the FEA, i.e., constant electric field magnitude per



(a) 3 dimensional view.



(b) 2 dimensional view.

FIG. 2. Finite-element model of a class IV flextensional transducer.

unit length of the piezoceramic. Hence, the value of the sound pressure in Fig. 5 has no practical meaning, but is good for comparison only. Results of the analysis of Fig. 5 are summarized in Fig. 6. For the calculation, the length of each ceramic element composing the stack was also changed to keep the initial ratio of the ceramic stack length to the semimajor axis length. An increase of the semimajor axis length caused the decrease in both the operation frequency and the sound pressure. With a larger horizontal dimension, the operation frequency of the transducer decreased, and this result agrees well with that in Ref. 5. The sound pressure is proportional to the product of the volume displacement amplitude and the square of the operation frequency of the transducer.[11] Since the amount of the decrease in the operation frequency was larger than that of the increase in the volume displacement amplitude, the resultant sound pressure was reduced.

Similar analyses were conducted to find the trends of the operation frequency and the maximum sound pressure change in relation to the variation of other structural vari-

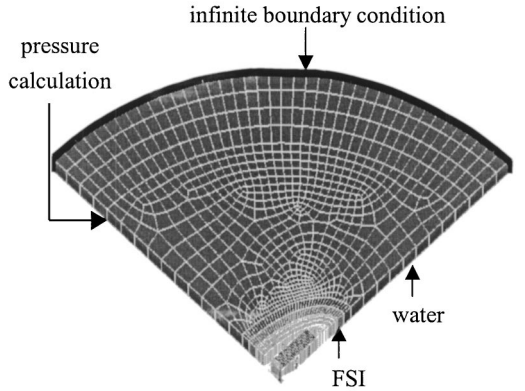
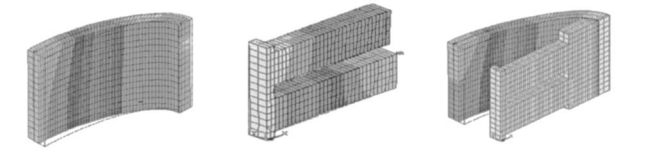


FIG. 3. Plane view of the whole FEA model.



(a) first flextensional mode of the shell

(b) first length extensional mode of the ceramic stack

(c) first vibration mode of the whole transducer

FIG. 4. Mode shape of the basic model.

TABLE I. Basic geometry of the flextensional transducer.

Semimajor axis length	200 mm
Thickness of the shell	40 mm
Height of the shell	173.8 mm
Minor axis length/major axis length	0.4
Ceramic stack length/semimajor axis length	0.6
Height of the end plate	40 mm
Thickness of the rubber mold	5 mm

ables. Figure 7 shows the influence of the shell thickness on the operation frequency as well as the maximum sound pressure. Both performance factors were observed to increase in proportion to the shell thickness. As reported earlier,[5] a thicker shell has a higher operation frequency and induces a smaller volume displacement amplitude.[20] The sound pressure generated by the transducer is dependent more on the operation frequency and thus increases in accordance with the shell thickness. Figure 8 shows the variation of the two performance factors in relation to the shell height. The volume of the transducer increases in proportion to the shell height. Hence, a higher shell causes a larger volume of the transducer, which results in a lower operation frequency. With an increase of the shell height, the height of the ceramic increases too, because its ratio to the shell height is kept constant at 0.8. A higher shell means higher ceramic stacks that can act as driving elements of the transducer, and thus supply more power to the transducer, which results in the increase of the sound pressure. Figure 9 shows the variation of the operation frequency and the sound-pressure level in relation to the ratio of the minor axis length to the major axis length. The ratio was varied by changing the minor axis length while keeping the major axis length constant. The operation frequency did not show much change with the variation of the ratio, which agrees well with previously reported results.[5] On the other hand, the sound pressure increased in response to the increase of the ratio. The edge of the ceramic stack is in contact with the interior surface of the shell. Hence, an increase of the minor axis length for a constant major axis length produces a wider piezoceramic stack that supplies more power to the transducer, thus creating a higher sound pressure. Finally, we increased the length of the ceramic stack while maintaining the length of the major axis constant, as shown in Fig. 10. For the constant major axis length, the increase of the ceramic stack length was made by reducing the length of the insert placed between the ceramic stack and the shell. In all cases, the ceramic stack contacted the interior surface of the shell. Hence, an increase of the ceramic stack length causes a decrease of the ceramic stack width, which means less power for the transducer over all and thus a lower sound pressure. Furthermore, as the length of the ceramic stack increases, in addition to the stack width, the width of the insulators, the insert, and the nodal plate is decreased as well. As a result, the interior of the shell becomes emptier, which leads to the decrease of the operation frequency as illustrated in Fig. 10. To summarize these results, shell thickness and major axis length had the largest influence on the operation frequency and sound-pressure level. The ratio of the minor axis length to the major axis length had little effect on the operation frequency but had a

TABLE II. Fixed dimensions and material properties of the parts in the transducer.

	Material	Young's modulus (Pa)	Density (kg/m^3)	Poisson's ratio
Shell, insert, nodal plate	Aluminum	68.9E9	2710	0.30
Mold	Rubber	3.0E7	1100	0.49
Insulator	Steatite	80.0E9	2900	0.25
End plate	Steel	210.0E9	7500	0.30
Ceramic stack	PZT-8	Ref. 19		

TABLE III. Variation range for the structural variables of a flextensional transducer.

Structural variables	Variation range
Semimajor axis length	150–250 mm
Thickness of the shell	20–60 mm
Height of the shell	142.5–205.0 mm
Minor axis length/major axis length	0.3–0.5
Ceramic stack length/semimajor axis length	0.5–0.7

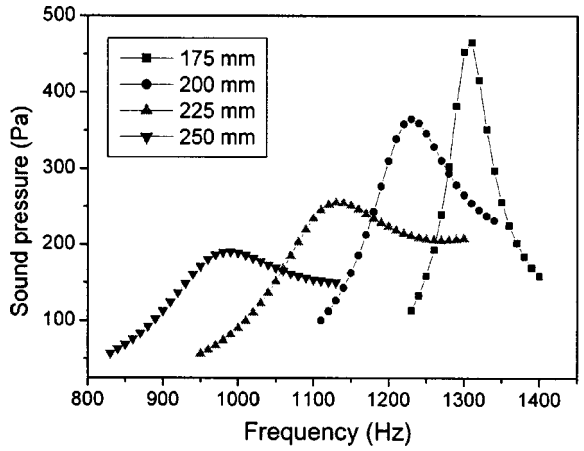


FIG. 5. FE harmonic analysis results showing the relationship between the transducer performance and the semimajor axis length.

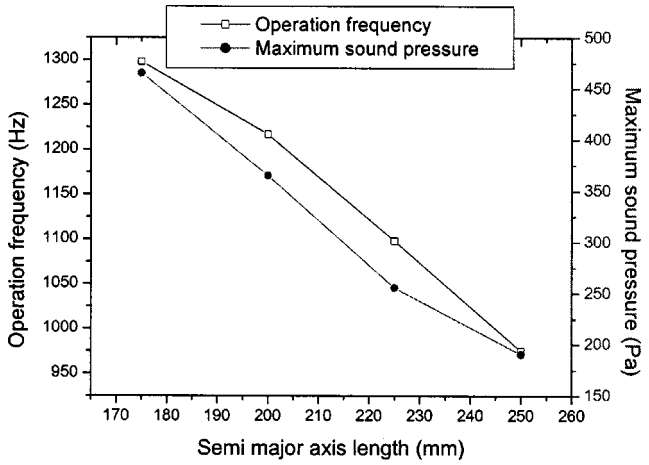


FIG. 6. Operation frequency and maximum sound pressure versus semimajor axis length.

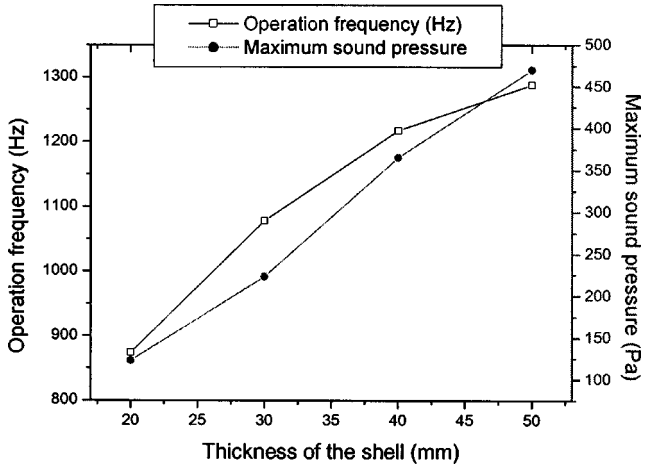


FIG. 7. Operation frequency and maximum sound pressure versus thickness of the shell.

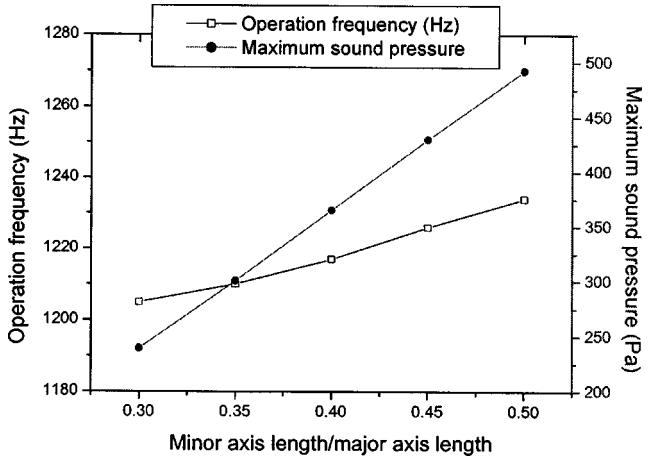


FIG. 9. Operation frequency and maximum sound pressure versus minor axis length/major axis length.

large effect on the sound-pressure level. The shell height and the ratio of the ceramic stack length to the semimajor axis length were not so influential as the other variables.

As previously stated, the purpose of the current study was to determine the optimal structure of a flextensional transducer that can provide the highest sound pressure at 1 kHz. Some of the variables showed consistent influence on the two performance factors, either positive or negative, while some others showed contradictory influences. In addition, the effects of the variables were not independent of but were cross coupled with each other. Hence, there should be a clever way to trade off the effects of each structural variable to achieve the optimal performance, and thus we employed the statistical multiple regression analysis technique and optimization technique for the purpose.

IV. STATISTICAL MULTIPLE REGRESSION ANALYSIS OF THE TRANSDUCER PERFORMANCE

Based on the results of the analysis discussed in Sec. III, we derived the functional forms of the operation frequency and the sound pressure of the transducer in terms of the structural variables. The derived functions included not only the individual but also all the cross-coupled effects of the structural variables. Formulation of the transducer performance was carried out by means of multiple regression analysis.[14,21]

In Sec. III, we analyzed 19 different combinations of structural variables to find the variation trends of the transducer performance. However, those 19 cases were not enough to compose the transducer performance as a function of the structural variables. Moreover, analysis of the structural effects through the FEA for all possible combinations of the variables was very time consuming, and was almost prohibitive for our flextensional transducer that has many structural variables. Hence, we employed the response surface design method to carry out the additional finite-element analysis in a systematic manner at the expense of least resources.[13] We employed the design method of 2^n experiments (n=number of the structural variable). Table IV shows the normalization factors established for each of the structural variables following the design. The normalization factors 1 and -1 correspond to the maximum and the minimum value of each of the structural variables, respectively. The normalization factor of 0 means the basic geometry of the transducer. The combination of the variables corresponding to the cases in Table IV numbered $2^5=32$. However, through

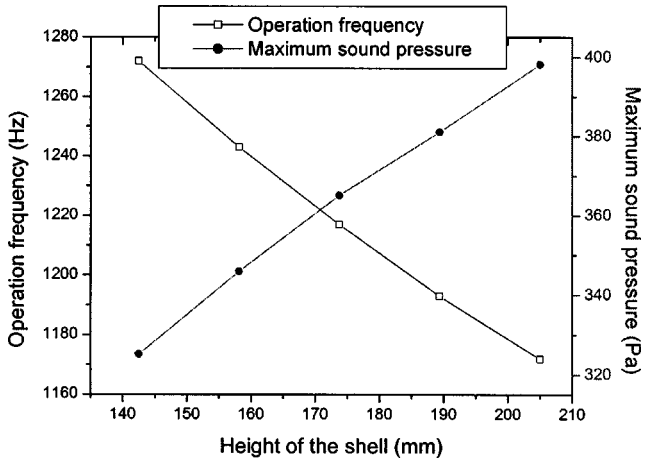


FIG. 8. Operation frequency and maximum sound pressure versus height of the shell.

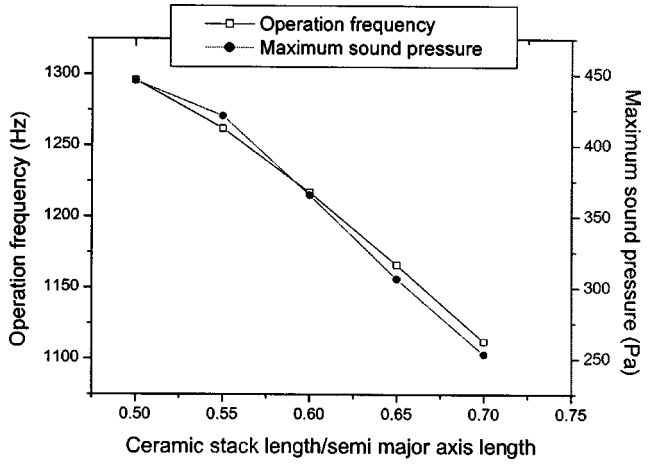


FIG. 10. Operation frequency and maximum sound pressure versus ceramic stack length/semimajor axis length.

TABLE IV. Normalization factors of the structural variables.

Normalization factors	Semimajor axis length (x_1) (mm)	Thickness of the shell (x_2) (mm)	Height of the shell (x_3) (mm)	Minor axis length/major axis length (x_4)	Ceramic stack length/semimajor axis length (x_5)
−1	150	20	142.5	0.3	0.5
0	200	40	173.8	0.4	0.6
1	250	60	205.0	0.5	0.7

preliminary calculations, we found that those 32 sets of data were insufficient for accurate formulation of the transducer performance. Thus, in addition to those extreme combinations, we analyzed the intermediate cases that corresponded to the normalization factors of ± 0.5 for each of the structural variables. The additional combinations of the variables with normalization factors of ± 0.5 also numbered $2^5=32$. Hence, in total, we carried out FEA for 83 $(19+32+32)$ different combinations of the structural variables.

For the statistical analysis of the FEA results to formulate the transducer performance, we developed a second-order multiple regression model that could consider the cross-coupled effects of the structural variables by the product terms of the variables. Regression models of more than third orders might give a better fit with the FEA results. However, the functional form would be too complicated to be practical.[14] A second-order regression model took the form in Eq. (1), and we had to find 37 regression coefficients to make the function coincide with the analysis data.

$$y = a_1 x_1^2 + \cdots + a_5 x_5^2 + b_1 x_1 + \cdots + b_5 x_5 + c_1 x_1 x_2 + \cdots + c_{10} x_4 x_5 + d_1 x_1 x_2 x_3 + \cdots + d_{10} x_3 x_4 x_5 + e_1 x_1 x_2 x_3 x_4 + \cdots + e_5 x_2 x_3 x_4 x_5 + f_1 x_1 x_2 x_3 x_4 x_5 + f_2, \quad (1)$$

where $a_1 - f_2$ are the 37 regression coefficients to be found, and $x_1 - x_5$ are the 5 structural variables in Table IV. The statistical analysis was made using the commercial software, SAS.[22] In the analysis, the coefficients having a level of significance smaller than 0.05 were removed because their influence on transducer performance was negligible. Through the analysis, the regression equation for the operation frequency, f_r, was determined as Eq. (2)

$$f_r = -14.63x_1^2 - 80.33x_2^2 + 7.76x_3^2 + 14.14x_4^2 + 2.46x_5^2 - 199.20x_1 + 205.61x_2 - 47.84x_3 + 19.20x_4 - 76.64x_5 + 52.74x_1x_2 + 11.15x_1x_3 + 27.06x_1x_4 - 12.99x_1x_5 - 9.99x_2x_3 - 48.54x_2x_4 + 6.73x_2x_5 - 4.91x_3x_4 - 11.70x_3x_5 + 6.98x_4x_5 + 12.44x_1x_2x_3 - 25.38x_1x_2x_5 + 1.95x_1x_3x_4 + 14.81x_1x_3x_5 + 5.22x_1x_4x_5 - 12.72x_2x_3x_5 - 14.62x_2x_4x_5 + 36.61x_3x_4x_5 + 20.16x_1x_2x_3x_5 + 23.82x_1x_2x_4x_5 - 37.86x_1x_3x_4x_5 + 35.90x_2x_3x_4x_5 - 38.21x_1x_2x_3x_4x_5 + 1200.85. \quad (2)$$

In Eq. (2), three regression coefficients exerting little influence on the operation frequency were removed, and there remained 34 coefficients. In the statistical regression analysis, the coefficient of determination describes the accuracy of the determined regression model representing the data to be fitted. For Eq. (2), the coefficient of determination was found to be 0.995, which confirmed that the equation described well the variation of the transducer operation frequency in relation to the structural variables. In the same manner, the regression equation for the sound pressure, P_r, was determined as Eq. (3)

$$P_r = 34.55x_1^2 + 6.72x_2^2 + 12.68x_4^2 - 186.50x_1 + 233.95x_2 + 29.81x_3 + 119.69x_4 - 82.14x_5 - 93.01x_1x_2 - 9.64x_1x_3 - 49.08x_1x_4 + 20.27x_1x_5 + 12.72x_2x_3 + 59.89x_2x_4 - 44.52x_2x_5 + 6.20x_3x_4 + 13.62x_4x_5 - 4.87x_1x_2x_3 - 14.85x_1x_2x_4 - 2.90x_1x_2x_5 - 17.93x_1x_4x_5 + 1.59x_2x_3x_4 + 13.92x_2x_4x_5 - 1.62x_3x_4x_5 + 2.11x_1x_2x_3x_5 - 15.36x_1x_2x_4x_5 - 2.97x_2x_3x_4x_5 - 1.87x_1x_2x_3x_4x_5 + 353.91. \quad (3)$$

Eight regression coefficients were removed due to their negligible level of significance, and there remained 29 coefficients. The coefficient of determination of Eq. (3) was 0.994, which confirmed again that the equation described well the variation of the sound pressure in relation to the structural variables of the transducer. As a confirmation of the reliability of Eqs. (2) and (3), trial calculations of transducer performance were carried out with arbitrary geometry of the transducer. The results represented in Table V show that the results of the regression analysis have some discrepancy with those from direct FEA. However, the two sets of results turned out to agree with each other within 2% difference at maximum for the operation frequency and about 3% difference at maximum for the sound pressure.

The functional forms of the operation frequency and sound pressure of the transducer are valid within the variation range in Table III of the structural variables. Once these functions become available, we can predict the transducer

TABLE V. Comparison of the analysis results with the derived transducer performance functions and the FEA.

	Operation frequency, f_r	Sound pressure, P_r
Eqs. (2) & (3)	924 Hz	212.9 Pa
FEA	906 Hz	207.0 Pa

performance for any combination of the variables within the range without carrying out the time-consuming FEA every time. Use of these functions in the optimization of the transducer structure will expedite the process with great efficiency.

V. OPTIMAL DESIGN OF A FLEXTENSIONAL TRANSDUCER STRUCTURE

With the functional forms of the two transducer performance factors derived in Sec. IV, we established a target function that should be minimized while satisfying several constraints created by the desired applications, and the minimization process was carried out with the SQP-PD method.

The main use of a flextensional transducer is to radiate an omnidirectional sound wave that has high power at a low frequency. To be omnidirectional, the largest dimension of the transducer should be smaller than one-third of its wavelength.[23] The transducer in this study is designed to work at 1 kHz for underwater applications. Hence, the maximum dimension of the transducer should be smaller than 500 mm, which corresponds to the maximum dimension of 250 mm for the symmetric model in Fig. 2. The maximum dimension in the Z direction in Fig. 2(a) is the shell height, which was limited to less than 205 mm, as in Table III. On the XY plane in Fig. 2(b), the maximum dimension l_x in the major axis direction is given as Eq. (4), which should be smaller than 250 mm for an omnidirectional radiation pattern

$$\begin{aligned} l_x = [&(\text{semimajor axis length} = 200 + 50x_1) \\ &+ (\text{thickness of the shell}/2 = 20 + 10x_2) \\ &+ (\text{thickness of the rubber mold} = 5)]. \end{aligned} \tag{4}$$

The length of the semimajor axis was measured from the center of the transducer to the neutral plane of the shell; hence, the l_x was counted with half of the shell thickness. The requirement that the l_x should be smaller than 250 mm worked as a dimensional constraint on the design of the transducer. The requirement that the transducer should operate with the center frequency of 1 kHz worked as another design constraint. The target function T was established as Eq. (5), with the two constraints described above as Eq. (6).

TABLE VI. Results of the optimization with the SQP-PD method.

Structural variables and performance	Results of optimization	Normalization factor
Semimajor axis length (x_1)	226 mm	0.52
Thickness of the shell (x_2)	38 mm	−0.1
Height of the shell (x_3)	205 mm	1.0
Minor axis length/major axis length (x_4)	0.5	1.0
Ceramic stack length/ semimajor axis length (x_5)	0.65	0.55
Operation frequency	1050 Hz	
Maximum dimension in the semimajor axis direction	250 mm	
Sound pressure	342.9 Pa	

Optimization of the transducer geometry was achieved by finding the combination of the structural variables that minimizes the target function while satisfying the two constraint equations

$$T(x_1, x_2, x_3, x_4, x_5) = |1/P_r|, \tag{5}$$

$$\begin{aligned} &\text{Subject to} \quad 950 \leqslant f_r \leqslant 1050 \\ &200 + 50x_1 + 20 + 10x_2 + 5 \leqslant 250. \end{aligned} \tag{6}$$

In the optimization, we allowed an error range of the achieved operation frequency of 5%, which meant that we allowed an operation frequency from 950 to 1050 Hz to be acceptable in the optimization process. For the minimization of the target function, the SQP-PD method was employed because of its high convergence rate among various kinds of constrained optimization methods.[15] Table VI shows results of the optimization. The dimensions in Table VI indicate the transducer geometry that can provide the highest sound-pressure level at the operation frequency of 1 kHz while satisfying the dimensional constraints. As expected in Sec. III, shell height and the ratio of the minor axis length to the major axis length had the maximum values allowed in Table III. The other three structural variables were found to have their optimal values inside the variation range. Validity of the optimized structure was checked through comparison with trial geometries of the transducer with other arbitrary dimensions. The comparisons were carried out rigorously with almost all the possible combinations of the five structural variables within the allowed variation range. Table VII shows some representative results, which confirm the superiority of the optimal geometry.

TABLE VII. Operation frequency and maximum sound pressure calculated with Eqs. (2) and (3) for the optimum geometry and comparative dimensions.

x_1	x_2	x_3	x_4	x_5	Operation frequency (Hz)	Maximum sound pressure (Pa)
0.52	−0.1	1.0	1.00	0.55	1050	342.9
0.50	−0.2	1.0	1.00	0.50	1037	327.8
0.50	−0.2	1.0	0.50	0.00	1038	303.5
0.50	−0.1	1.0	0.75	0.50	1040	320.9

VI. SUMMARY

In this paper, we proposed an optimal design method for a flextensional transducer. The design process consisted of the following three steps. First, through the FEA we analyzed the variation of the operation frequency and the sound pressure of the transducer in relation to its structural variables. A design method of 2^n experiments was employed to reduce the number of analysis cases. Second, through statistical multiple regression analysis of the results, we derived functional forms of the transducer performance that could consider the cross-coupled effects of the structural variables. Third, we determined the optimal geometry of the transducer that had the highest sound pressure at the desired operation frequency through the optimization with the SQP-PD method of a target function composed of the transducer performance. The proposed method was illustrated with a sample design of a transducer working at 1 kHz, and the validity of the optimized result was confirmed through comparison of the optimal performance with that of trial transducers of different geometries.

The performance of an acoustic transducer is determined by the effects of many structural variables. Most influences of these structural variables are not linearly independent of each other. To achieve the optimal performance of an acoustic transducer, we must consider the cross-coupled effects of its structural variables. However, analysis of all the effects through the FEA is very time consuming, and is almost prohibitive for a transducer with many structural variables like the flextensional transducer considered in this paper. The optimal design method proposed could reflect all the cross-coupled effects of multiple structural variables, and could determine the detailed geometry of the transducer with great efficiency and rapidity. However, when the nonlinear interactions of the variables have strong effects on the transducer performance, it is desirable to carry out more rigorous validation of the regression equation before real application of the design method. The proposed technique is so general that it can also be applied to any other type of acoustic transducers.

ACKNOWLEDGMENT

This work was supported by the Brain Korea 21 Project in 2002.

[1] S. C. Butler, J. L. Butler, A. L. Butler, and G. H. Cavanagh, ''A low-frequency directional flextensional transducer and line array,'' J. Acoust. Soc. Am. **102**(1), 308–314 (1997).
[2] K. D. Rolt, ''History of the flextensional electroacoustic transducer,'' J. Acoust. Soc. Am. **87**(3), 1340–1348 (1990).
[3] G. Bromfield, ''Class IV flextensional transducers,'' in *Power Transducers for Sonics and Ultrasonics*, edited by B. Hamonic, O. B. Wilson, and J. N. Decarpigny (Springer, Berlin, 1990), Chap. 5, pp. 48–59.
[4] D. Boucher, ''Trends and problems in low frequency sonar projectors design,'' in *Power Sonics and Ultrasonic Transducers Design*, edited by B. Hamonic and J. N. Decarpigny (Springer, Berlin, 1987), Chap. 5, pp. 100–120.
[5] J. R. Oswin and A. Turner, ''Design limitations of aluminium shell, class IV flextensional transducers,'' Proc. Inst. Acoust. **6**(3), 94–101 (1984).
[6] M. B. Moffett, A. E. Clark, M. Wun-Fogle, J. Linberg, J. P. Teter, and E. A. McLaughlin, ''Characterization of Terfenol-D for magnetostrictive transducers,'' J. Acoust. Soc. Am. **89**(3), 1448–1455 (1991).
[7] C. L. Hom and N. Shankar, ''Modeling nonlinearity in electrostrictive sonar transducers,'' J. Acoust. Soc. Am. **104**(4), 1903–1913 (1998).
[8] R. J. Meyer, Jr., A. Dogan, C. Yoon, S. M. Pilgrim, and R. E. Newnham, ''Displacement amplification of electroactive materials using the cymbal flextensional transducer,'' Sens. Actuators A **87**(3), 157–162 (2001).
[9] R. Hanel, A. Mues, and R. Sobotta, ''Description of ultrasound transducers through wave parameters,'' Ultrasonics **34**, 159–162 (1996).
[10] P. Dufourcq, J. Adda, M. Letiche, and E. Sernit, ''Transducers for great depths,'' in *Power Transducers for Sonics and Ultrasonics*, edited by B. Hamonic, O. B. Wilson, and J. N. Decarpigny (Springer, Berlin, 1990), Chap. 7, pp. 75–85.
[11] J. R. Oswin and J. Dunn, ''Frequency, power and depth performance of class IV flextensional transducers,'' in *Power Sonics and Ultrasonic Transducers Design*, edited by B. Hamonic and J. N. Decarpigny (Springer, Berlin, 1987), Chap. 6, pp. 121–133.
[12] M. Romdhane, A. Gadri, F. Contamine, C. Gourdon, and G. Casamatta, ''Experimental study of the ultrasound attenuation in chemical reactors,'' Ultrason. Sonochem. **4**(3), 235–243 (1997).
[13] R. O. Kuehl, *Design of Experiments: Statistical Principles of Research Design and Analysis* (Duxbury, Pacific Grove, CA, 2000).
[14] R. J. Freund and W. J. Wilson, *Regression Analysis: Statistical Modeling of a Response Variable* (Academic, San Diego, 1998).
[15] A. D. Belegudu and T. R. Chandrupatla, *Optimization Concepts and Applications in Engineering* (Prentice-Hall, Englewood Cliffs, NJ, 1999), pp. 141–221.
[16] D. T. I. Francis, J. R. Oswin, and P. C. Macey, ''Comparing FE/BE models with measurement: Flextensional transducers,'' Proc. Inst. Acoust. **18**(10), 31–40 (1996).
[17] J. C. Debus, J. N. Decarpigny, and B. Hamonic, ''Analysis of a class IV flextensional transducer using Piece-Part equivalent circuit models,'' in *Transducers for Sonics and Ultrasonics*, edited by M. D. McCollum, B. Hamonic, and O. B. Wilson (Technomic, PA, 1992), Chap. 16, pp. 181–197.
[18] Y. W. Lam, ''Mathematical model of a class IV flextensional transducer and its numerical solution,'' Appl. Acoust. **36**, 123–144 (1992).
[19] O. B. Wilson, *Introduction to Theory and Design of Sonar Transducers* (Peninsula, Los Altos, CA, 1988), pp. 65–88.
[20] K. Onitsuka, A. Dogan, Q. Xu, S. Ysohikawa, and R. E. Newnham, ''Design optimization for metal–ceramic composite actuator, 'Moonie','' Ferroelectrics **156**, 37–42 (1994).
[21] R. E. McKeighen, ''Optimization of broadband transducer designs by use of statistical design of experiments,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **UFFC-43**(1), 63–70 (1996).
[22] STRATEGIC APPLICATIONS SOFTWARE, ver. 8.1 (SAS Institute, Inc., Cary, North Carolina, 2000).
[23] L. L. Beranek, *Acoustics* (American Institute of Physics, New York, 1988), pp. 91–115.

Outer hair cell piezoelectricity: Frequency response enhancement and resonance behavior

Erik K. Weitzel
Bobby R. Alford Department of Otorhinolaryngology and Communicative Sciences, Baylor College of Medicine, Houston, Texas 77030

Ron Tasker
TASI Technical Software, Inc., 444 Frontenac Street, Kingston, Ontario K7L 3T4, Canada

William E. Brownell[a)]
Bobby R. Alford Department of Otorhinolaryngology and Communicative Sciences, Baylor College of Medicine, Houston, Texas 77030

(Received 5 February 2003; revised 23 May 2003; accepted 9 June 2003)

Stretching or compressing an outer hair cell alters its membrane potential and, conversely, changing the electrical potential alters its length. This bi-directional energy conversion takes place in the cell's lateral wall and resembles the direct and converse piezoelectric effects both qualitatively and quantitatively. A piezoelectric model of the lateral wall has been developed that is based on the electrical and material parameters of the lateral wall. An equivalent circuit for the outer hair cell that includes piezoelectricity shows a greater admittance at high frequencies than one containing only membrane resistance and capacitance. The model also predicts resonance at ultrasonic frequencies that is inversely proportional to cell length. These features suggest all mammals use outer hair cell piezoelectricity to support the high-frequency receptor potentials that drive electromotility. It is also possible that members of some mammalian orders use outer hair cell piezoelectric resonance in detecting species-specific vocalizations.
[DOI: 10.1121/1.1596172]

PACS numbers: 43.38.Fx, 43.64.Bt, 43.64.Ld [WPS]

I. INTRODUCTION

Early in the 1880s, Jacques and Pierre Curie reported that the application of mechanical stress to one of a variety of crystalline materials produced electricity and coined the term piezoelectricity (from the Greek, *piezein*—to press). After considering the thermodynamics of the process, Gabriel Lippmann concluded that applying an electric field across the crystals should produce mechanical stress. The Curie brothers quickly confirmed the converse piezoelectric effect predicted by Lippmann (Cady, 1946). Crystalline piezoelectric materials have been used in applications that range from marine sonar to mechanical actuators to electronic delay lines. Noncrystalline, biological piezoelectricity has been observed in bone (Fukada and Yasuda, 1957) and ligament (Korostoff, 1977; Fukada, 1982).

Outer hair cell (OHC) electromotility (Brownell *et al.*, 1985) is a direct conversion of membrane potential to cell length change and is functionally equivalent to the converse piezoelectric effect. The mechanism responsible for electromotility resides in the OHC lateral wall, a nanoscale composite of membranes and cytoskeletal proteins (Brownell *et al.*, 2001). The OHC lateral wall also supports a mechano-electrical conversion (the direct piezoelectric effect) (Gale and Ashmore, 1994; Zhao and Santos-Sacchi, 1999), which further strengthens the view that the lateral wall is piezoelectric. Several theoretical treatments of the OHC and its role in hearing have assumed the OHC to be piezoelectric (Mountain and Hubbard, 1994; Tolomeo and Steele, 1995; Spector *et al.*, 1998, 1999; Raphael *et al.*, 2000; Iwasa, 2001; Spector *et al.*, 2002). The models all infer or assume Maxwell reciprocity (Cady, 1946) in which the coefficients for the direct and converse effects are equal. Recent experimental evidence provides quantitative support for this thermodynamic requirement for piezoelectricity (Dong *et al.*, 2002).

Electomotility makes a major contribution to the sensitivity and frequency selectivity of mammalian hearing (Brownell *et al.*, 2001; Liberman *et al.*, 2002). The motor mechanism responsible is membrane based, does not depend on calcium or cellular stores of ATP, and is capable of generating mechanical force at frequencies >50 kHz (Brownell and Kachar, 1986; Kachar *et al.*, 1986; Frank *et al.*, 1999; Brownell *et al.*, 2001). The functional significance of electromotility has been questioned because an equivalent circuit analysis based on conventional membrane resistance and capacitance predicts a low-pass frequency response that would limit the membrane voltage and therefore force production only to low frequencies (Hudspeth and Logothetis, 2000). We show that the inclusion of piezoelectricity into the equivalent circuit of the outer hair cell pushes the corner frequency to higher values. In addition, piezoelectric materials can be driven into resonance and we present an analysis predicting resonance in OHCs in the ultrasonic range. The predicted values are, however, compatible with those used for echolocation in chiroptera (bats) and cetaceans (whales and dolphins).

[a)]Author to whom correspondence should be addressed at One Baylor Plaza NA 505, Houston, TX 77030. Electronic mail: brownell@bcm.tmc.edu

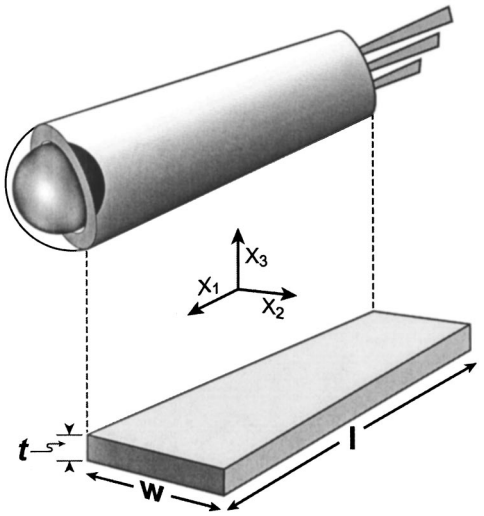


FIG. 1. Schematic of cylindrical OHC lateral wall (above) and its LTE representation (below).

II. METHODS

Lateral wall piezoelectricity is modeled for two different representations of the outer hair cell lateral wall, a rectangular slab and a cylindrical shell. The piezoelectric and material properties are assumed to be homogenous for both geometries. The rectangular slab is known as a length-thickness extension (LTE) resonator in the piezoelectric literature. Most of our analysis is developed for the LTE model because its properties are the most thoroughly studied. The cylindrical shell more closely matches the geometry of the OHC lateral wall and it is called a radial-poled cylinder (RPC) resonator.

III. RESULTS

For the purpose of analysis, the OHC lateral wall is split along its length and rolled out to form a rectangular slab detached from other cellular elements (Fig. 1). The slab is laid out along three Cartesian axes. The circumference of the cylinder is represented by the width (w) of the slab, parallel to x_2; the cell length (l) and wall thickness (t) are parallel to x_1 and x_3, respectively. The lateral wall is a composite structure composed of the plasma membrane, cortical lattice, and subsurface cisterna. OHC electromotility involves electromechanical coupling between all three layers (Brownell *et al.*, 2001) and we assume a thickness of 100 nm (seen in electron micrographs). An electric field across the lateral wall defines the direction of poling or net polarization across the thickness (t) of the slab (x_3 direction) and the resonance mode of interest induced by this field is along the length of the cell (x_1 direction). This is a LTE resonator the properties of which are fully described in the IEEE standard (Rosen *et al.*, 1992). The pertinent constitutive piezoelectric equations are

$$S = s^E T - dE, \quad D = dT + \varepsilon_0 \varepsilon^T E, \tag{1}$$

where S is the strain, D is the electric polarization (or displacement), s^E is the elastic compliance at constant electric field, ε_0 is the permittivity of free space, ε^T is the permittivity coefficient of the slab at constant stress, T is the stress, E is the electric field, and d is the piezoelectric constant of the slab. Equation (1) is a set of tensor equations, having electrical and mechanical properties in three dimensions where the mechanical properties include shear. The off-diagonal coefficients (d) are the same, as required by Maxwell reciprocity, and, specifically, $d_{31} = d_{13}$ in the complete LTE tensor representation (Rosen *et al.*, 1992). For a thin plate, $t \ll l$, the LTE resonator can be represented by a plane electrical wave propagating in the thickness direction coupled to a plane mechanical wave propagating in the length direction and Eq. (1) becomes

$$S_1 = s_{11}^E T_1 - d_{31} E_3, \quad D_3 = d_{31} T_1 + \varepsilon_0 \varepsilon_{33}^T E_3. \tag{2}$$

For unclamped standing waves in the length-direction, stress is zero at the ends of the specimen and S_1 becomes a function only E_3. The two parts of Eq. (2) are used to solve for D_3. Current as a function of frequency (f) is determined by integrating D_3 with respect to time. Dividing by voltage, we obtain the admittance of the specimen:

$$Y(f) = i \frac{2\pi f l w}{t} \left(\varepsilon_0 \varepsilon_{33}^T - \frac{d_{31}^2}{s_{11}^E} \right) + i \frac{2 w d_{31}^2}{t s_{11}^E \sqrt{\rho s_{11}^E}} \tan\left(\pi f l \sqrt{\rho s_{11}^E} \right). \tag{3}$$

See Table I for definitions of the symbols. The first term in $Y(f)$ after expanding the right side, reveals a conventional dependence on the E field while the 2nd and 3rd terms result from the dependence of S on E. Note that the quantity $\sqrt{\rho s_{11}^E}$ is the inverse of the speed of sound in the lateral wall.

Experimental values for the coefficients (Table I) are obtained from the literature (Halter *et al.*, 1997; Spector *et al.*, 1998, 1999). There have been no direct measures of the elastic compliance (s_{11}^E) and the piezoelectric coefficient (d_{31}^2) for the lateral wall. The first may be calculated from a one-dimensional determination of the axial Young's modulus (Y_{11}) for the OHC modeled as a spring with spring constant (k_1). A range of values for the axial Young's modulus has been reported and we choose one that is near the mean. It

TABLE I. Coefficient values. LW—lateral wall; PE—piezoelectric.

Coefficient	Symbol	Value	Unit
Length	l	20,40,60,80	μm
Radius	r	4.5	μm
Width (circumference)	w	28.5	μm
Thickness (LW & PE plate)	t	100[a]	nm
Density	ρ	1[a]	g/ml
PE plate permittivity	$\varepsilon_0 \varepsilon_{33}^T$	5.0×10^{-11}	C^2/Nm^2
PE plate elastic compliance	s_{11}^E	3.1×10^{-5}	m^2/N
PE constant	d_{31}^2	6.8×10^{-16}	C^2/N^2
LW resistance	R_{pm}	100	Ωm^2
LW capacitance	C_{pm}	1.7×10^{-2}	F/m^2
Basal membrane resistance	R_{ba}	1	Ωm^2
Basal membrane capacitance	C_{ba}	1.7×10^{-2}	F/m^2
Apical area (cuticular plate)	A_{cp}	8.2×10^{-11}	m^2
Apical membrane resistance	R_{ap}	5.0×10^{-4}	Ωm^2
Apical membrane capacitance	C_{ap}	1.7×10^{-2}	F/m^2

[a]Estimated—this paper.

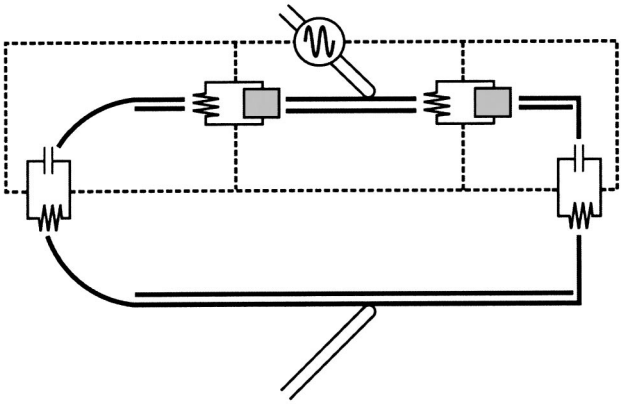


FIG. 2. Equivalent circuit diagram for microchamber recording conditions. Gray box impedance in the lateral wall represents either a piezoelectric element [from Eq. (3)] or pure capacitance for the conventional RC analysis.

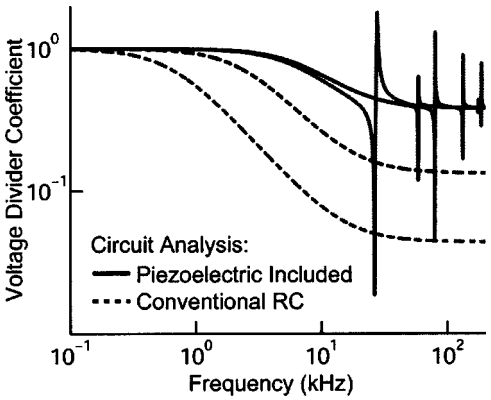


FIG. 3. Equivalent circuit analysis for 20 and 80 μm OHCs with 1% of the cell in the microchamber. The solid and dashed gray lines with the lower cutoff and resonance frequency represent the 80 μm cell.

was obtained by compressing the OHC with a calibrated cantilever without controlling the membrane potential (Holley and Ashmore, 1988). The elastic compliance of a hollow cylinder of length (l), width (w), radius (r), and thickness (t) is

$$s_{11}^E = Y_{11}^{-1} = \frac{\text{strain}}{\text{stress}} = \frac{\Delta L_1/L_1}{F_1/A} = \frac{\Delta L_1/L_1}{k_1 \Delta L_1/A} = \frac{A}{k_1 L_1}, \quad (4)$$

where $A = \pi(r)^2 - \pi(r-t)^2 \cong 2\pi rt = wt$ and $L_1 = l$.

The piezoelectric coefficient d_{31}^2 is calculated from

$$d_{31} = \frac{\text{strain}_1}{\text{ElectricField}_3} = \frac{\Delta L_1/L_1}{V_3/t} = \frac{S_1}{V_3} t. \quad (5)$$

The values for S_1/V_3 (where V_3 is the potential difference across the wall thickness) are available from the literature (Tolomeo and Steele, 1995).

An equivalent circuit of the OHC that included piezoelectricity was developed to evaluate its influence on the membrane potential. A model of current flow across OHC membranes for a cell held in a microchamber (Dallos and Evans, 1995) was adapted. The model retained the two-dimensional radially symmetric RC circuit of the original, with a piezoelectric element (Sherrit *et al.*, 1997) added in parallel to the lateral wall membrane resistance (Fig. 2). The admittance of the element is that defined for the LTE model in Eq. (3). The element is added in parallel because the piezoelectric element is assumed to be an ideal material with no losses so that there would be no current at DC. Two simulations were performed (1) with piezoelectricity and (2) with a capacitor (the membrane capacitance alone).

The results of these simulations are seen in Fig. 3 in which the voltage-divider equivalent impedance is plotted against frequency. Piezoelectricity pushes the corner frequency to a higher value and diminishes the magnitude of dampening. Figure 3 also provides a comparison between cells of different length (20 vs 80 μm). Resonance and corner frequencies are increased in the shorter cell while the overall electrical behavior is similar.

Based on the IEEE-defined formula for the series resonance frequency $f_S = (1/2l)\sqrt{1/\rho s_{11}^E}$, the first resonance frequency for the piezoelectric slab (80 μm) is predicted at 37 kHz. An iterative solution of

$$\frac{(k_{31}^l)^2}{1-(k_{31}^l)^2} = \frac{\pi}{2}\frac{f_p}{f_S}\tan\left(\frac{\pi}{2}\frac{\Delta f}{f_S}\right)$$

predicts the anti-resonance frequency (f_p) at 46 kHz, where $(k_{31}^l)^2 = d_{31}^2/\varepsilon_{33}^T s_{11}^E$ is the electromechanical coupling constant. Additional confirmation of OHC piezoelectricity would come from comparing impedance and admittance in isolated cells to confirm a shift between resonance (f_s) and antiresonance (f_p). Resonance frequencies increase with decreasing cell length (Fig. 4). This is intuitive as resonance frequency is inversely proportional to length for most materials. An inverse relationship was also found between resonance frequency and slab thickness (Fig. 5).

A radial-poled cylinder (RPC) resonator (Ebenezer and Sujatha, 1997) was also examined. It allowed an assessment of not considering the contribution of S_2 resonance mode in the development of the LTE model. The RPC model consists of a cylindrical shell where the diameter of the cylinder is much greater than the thickness of the shell and the electric field acts across the shell thickness. The d_{31} coefficient (as defined in the LTE mode) couples to the circumferential di-

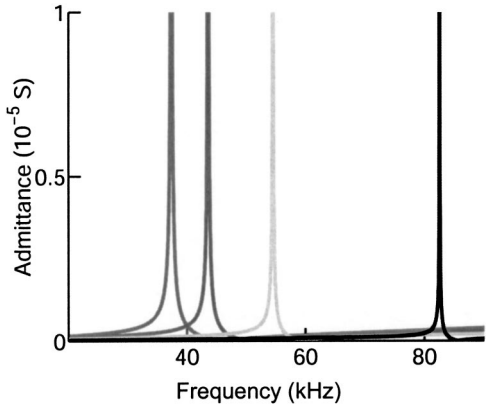


FIG. 4. Admittance resonance peaks for LTE model of different lengths. The frequency of the resonance peaks are higher for shorter cells. The cell lengths are 80, 60, 40, and 20 μm proceeding from left to right.

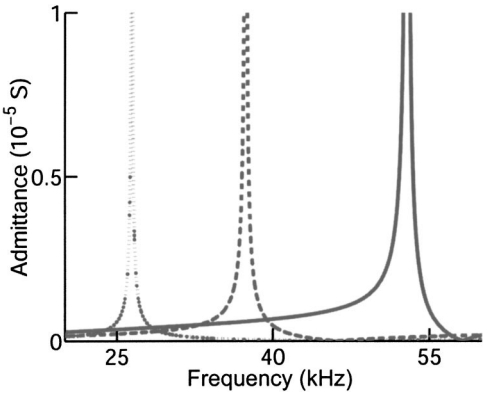


FIG. 5. Effect of lateral wall thickness on admittance resonance spectrum for an LTE model of an 80-μm-long cell. All other parameters as in Fig. 4. Resonance frequency increases with decreasing thickness. The wall thickness is 200, 100, and 50 nm proceeding from left to right.

rection causing a displacement (described by s_{11}^E) manifesting itself as a hoop or breathing mode. This then couples to the length of the cylinder through s_{12}^E to result in the radial-poled cylinder resonator. The resonance spectra of different length radial-poled cylinders were nearly identical to those of the rectangular slab model when l (cell length)$>$60 μm. The resonance of shorter cylinders was less than that of the same length LTE model (Fig. 6), showing that inclusion of an S_2 resonance mode becomes more important in the LTE model as l becomes shorter.

IV. DISCUSSION

An unresolved challenge to the OHC's role as the cochlear amplifier has been the fact that a conventional circuit analysis based on passive resistive and capacitive elements predicts a low-pass behavior as with the dashed curves in Fig. 3 (Hudspeth and Logothetis, 2000). This should restrict OHC electromotility to low frequencies below a corner frequency determined by the length of the cell. A thermodynamic approach suggests a piezoelectric augmentation of the OHC receptor potential that results from the OHC deformation that occurs during normal cochlear vibrations (Spector *et al.*, 2002, 2003). The present results, based on an

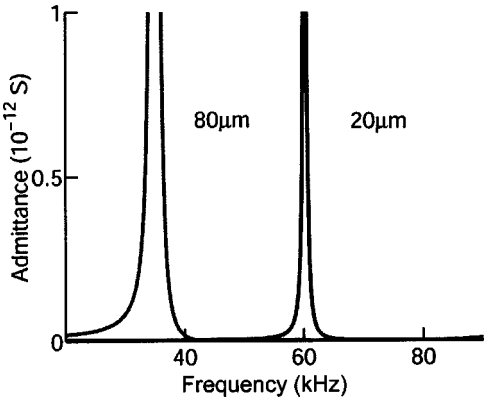


FIG. 6. Radial poled cylinder (RPC) model. All parameters are the same as those used in the LTE model. The elastic compliance in the circumferential direction was chosen to be the same as that for the axial direction. Resonance frequency is similar to the LTE at 80 μm length, but at 20 μm length the resonance is not as high a frequency as that of the LTE model.

engineering analysis of the lateral wall, reveal that OHC piezoelectricity would also extend the frequency range of stereocilia mediated sensory receptor currents. Electromotility is mediated by the transmembrane receptor potential. Future cochlear modeling should consider the contributions to the OHC receptor potential from both ionic currents and piezoelectric modulation of the membrane potential by acoustically driven cell deformation.

The extended frequency range predicted from our model is consistent with the corner frequencies measured for mechanically unloaded OHCs when stimulated in the microchamber configuration (Dallos and Evans, 1995; Frank *et al.*, 1999). Even though a mechanical resonant peak has not been observed in isolated OHCs, Frank *et al.* (1999) attributed the electrically induced displacement response, with its elevated cutoff frequency, to the presence of a second-order resonance. This conclusion was based on fitting the displacement response, in both amplitude and phase, with that of an overdamped second-order resonant system, in which the resonant frequency was found to depend on cell length. Their data required the presence of a length-dependent resonance similar to that shown in Fig. 4.

The OHC is mechanically loaded in the fluid filled organ of Corti. It is firmly anchored to supporting cells by tight junctions in the reticular lamina. The tectorial and basilar membranes introduce additional mechanical loads. The *in vivo* OHC is therefore a viscous-damped mechanically loaded resonator in contrast to the unloaded or free resonator in our analysis. Radiation damping of the resonance would be expected to result in a broadening and weakening of the impedance peak as a function of frequency. The observation of a broad, ultrasonic mechanical resonance in electrically evoked basilar membrane vibrations near the round window of the guinea pig inner ear (Grosh *et al.*, 2003) is consistent with this expectation. The OHCs near the round window are all short and close to the same length. By inference they should have similar resonance behavior. Electrical stimulation around this frequency would result in a summed response that would be broader than that expected in an unloaded OHC. Many rodents display ultrasonic vocalizations during mating (White *et al.*, 1998). The broad electrically evoked ultrasonic vibrations observed in the guinea pig ear (Grosh *et al.*, 2003) contain these mating frequencies.

Outer hair cells appeared with mammals over 200 million years ago. The OHCs of the small rodentlike early mammals enabled them to hear frequencies above 50 kHz. They were the basis of a cochlear amplifier that counteracted viscous damping and refined the passive mechanical filtering of the organ of Corti (Brownell *et al.*, 2001). Hair cell resonance would not benefit hearing in most mammals and it is of interest that the predicted resonance frequencies are generally beyond the frequency range that OHCs at a specific location on the basilar membrane would operate. In addition, the impedance mismatch between the OHCs and the organ of Corti is so severe that any resonance would be expected to remain confined within the cell and have little impact on the vibrations of the organ of Corti. The importance of OHC biological piezoelectricity for all mammals appears to be its ability to improve the high-frequency electrical response and

allow the cochlear amplifier to work throughout the mammalian frequency range.

The first bats appeared over 60 million years ago with an elegant echolocation mechanism that allowed them to navigate in a nocturnal environment. An additional resonance has been invoked to explain their enhanced hearing sensitivity at the echolocation frequencies (Kossl and Russell, 1995; Russell and Kossl, 1999). Bats may have co-opted OHC piezoelectric resonance to achieve this. Some 10 million years later cetaceans adopted a remarkably similar specialization (Ketten, 1997). The organ of Corti in echolocating members of the orders *chiroptera* and *cetacea* is greatly expanded at the region devoted to the echolocation frequencies. The OHCs in the region are short and of uniform length so that the reticular lamina is parallel to the basilar membrane. This organization would allow OHC resonance to be summed, overcoming the impedance mismatch that normally obscures the resonance. While echolocating mammals may be the only ones to utilize the cellular resonance, the piezoelectric mechanism that underlies it helps to support high-frequency receptor potentials throughout the entire class *mammalia*.

ACKNOWLEDGMENTS

Research supported by National Science Foundation Nanotechnology Initiative Research Grant No. BES-9871994 and Public Health Service Research Grant No. DC 00354 from NIDCD. The authors would like to thank Dr. Grosh, Dr. Gummer, Dr. Oghalai, Dr. Rabbitt, Dr. Raphael, and Dr. Spector for helpful comments and Nainesh Gandhi for his assistance with data entry in the cylindrical model.

Brownell, W. E., and Kachar, B. (**1986**). ''Outer hair cell motility: A possible electro-kinetic mechanism,'' in *Peripheral Auditory Mechanisms*, edited J. B. Allen, J. L. Hall, A. E. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, Berlin), pp. 369–376.

Brownell, W. E., Bader, C. R., Bertrand, D., and de Ribaupierre, Y. (**1985**). ''Evoked mechanical responses of isolated cochlear outer hair cells,'' Science **227**, 194–196.

Brownell, W. E., Spector, A. A., Raphael, R. M., and Popel, A. S. (**2001**). ''Micro- and Nanomechanics of the Cochlear Outer Hair Cell,'' Annu. Rev. Biomed. Eng. **3**, 169–194.

Cady, W. G. (**1946**). *Piezoelectricity* (McGraw-Hill, New York).

Dallos, P., and Evans, B. N. (**1995**). ''High-frequency motility of outer hair cells and the cochlear amplifier,'' Science **267**, 2006–2009.

Dong, X. X., Ospeck, M., and Iwasa, K. H. (**2002**). ''Piezoelectric reciprocal relationship of the membrane motor in the cochlear outer hair cell,'' Biophys. J. **82**, 1254–1259.

Ebenezer, D. D., and Sujatha, A. J. (**1997**). ''New methods to characterize radially polarized piezoelectric ceramic cylindrical shells of finite length,'' J. Acoust. Soc. Am. **102**, 1540–1548.

Frank, G., Hemmert, W., and Gummer, A. W. (**1999**). ''Limiting dynamics of high-frequency electromechanical transduction of outer hair cells,'' Proc. Natl. Acad. Sci. U.S.A. **96**, 4420–4425.

Fukada, E. (**1982**). ''Electrical phenomena in biorheology,'' Biorheology **19**, 15–27.

Fukada, E., and Yasuda, I. (**1957**). ''On the piezoelectric effect of bone,'' J. Phys. Soc. Jpn. **12**, 1158–1162.

Gale, J. E., and Ashmore, J. F. (**1994**). ''Charge displacement induced by rapid stretch in the basolateral membrane of the guinea-pig outer hair cell,'' Proc. R. Soc. London, Ser. B **255**, 243–249.

Grosh, K., Zheng, J., deBoer, E., and Nuttall, A. L. (**2003**). ''High frequency electromotile responses in the cochlea,'' personal communication.

Halter, J. A., Kruger, R. P., Yium, M. J., and Brownell, W. E. (**1997**). ''The influence of the subsurface cisterna on the electrical properties of the outer hair cell,'' NeuroReport **8**, 2517–2521.

Holley, M. C., and Ashmore, J. F. (**1988**). ''A cytoskeletal spring in cochlear outer hair cells,'' Nature (London) **335**, 635–637.

Hudspeth, A. J., and Logothetis, N. K. (**2000**). ''Sensory systems,'' Curr. Opin. Neurobiol. **10**, 631–641.

Iwasa, K. H. (**2001**). ''A two-state piezoelectric model for outer hair cell motility,'' Biophys. J. **81**, 2495–2506.

Kachar, B., Brownell, W. E., Altschuler, R. A., and Fex, J. (**1986**). ''Electrokinetic shape changes of cochlear outer hair cells,'' Nature (London) **322**, 365–368.

Ketten, D. R. (**1997**). ''Structure and function in whale ears,'' Bioacoustics **8**, 103–135.

Korostoff, E. (**1977**). ''Stress generated potentials in bone: relationship to piezoelectricity of collagen,'' J. Biomech. **10**, 41–44.

Kossl, M., and Russell, I. J. (**1995**). ''Basilar membrane resonance in the cochlea of the mustached bat,'' Proc. Natl. Acad. Sci. U.S.A. **92**, 276–279.

Liberman, M. C., Gao, J., He, D. Z., Wu, X., Jia, S., and Zuo, J. (**2002**). ''Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier,'' Nature (London) **419**, 300–304.

Mountain, D. C., and Hubbard, A. E. (**1994**). ''A piezoelectric model of outer hair cell function,'' J. Acoust. Soc. Am. **95**, 350–354.

Raphael, R. M., Popel, A. S., and Brownell, W. E. (**2000**). ''A membrane bending model of outer hair cell electromotility,'' Biophys. J. **78**, 2844–2862.

Rosen, C. Z., Hiremath, B. V., and Newnham, R. (**1992**). ''IEEE standard on piezoelectricity (ANSI/IEEE Standard 176-1987 1988),'' *Piezoelectricity* (American Institute of Physics, New York), pp. 227–228.

Russell, I. J., and Kossl, M. (**1999**). ''Micromechanical responses to tones in the auditory fovea of the greater mustached bat's cochlea,'' J. Neurophysiol. **82**, 676–686.

Sherrit, S., Widerick, H. D., Mukherjee, B. K., and Sayer, M. (**1997**). ''An accurate equivalent circuit for the unloaded piezoelectric vibrator in the thickness mode,'' J. Phys. D **30**, 2354–2363.

Spector, A. A., Brownell, W. E., and Popel, A. S. (**1998**). ''Estimation of elastic moduli and bending stiffness of the anisotropic outer hair cell wall,'' J. Acoust. Soc. Am. **103**, 1007–1011.

Spector, A. A., Brownell, W. E., and Popel, A. S. (**1999**). ''Nonlinear active force generation by cochlear outer hair cell,'' J. Acoust. Soc. Am. **105**, 2414–2420.

Spector, A. A., Brownell, W. E., and Popel, A. S. (**2003**). ''Effect of outer hair cell membrane piezoelectric properties on the receptor potential under high frequency conditions,'' J. Acoust. Soc. Am. **113**, 453–461.

Spector, A. A., Popel, A. S., and Brownell, W. E. (**2002**). ''Piezoelectric properties enhance outer hair cell high-frequency response,'' in *Biophysics of the Cochlea From Molecule to Model*, edited by A. W. Gummer (World Scientific, Singapore), pp. 152–160.

Tolomeo, J. A., and Steele, C. R. (**1995**). ''Orthotropic piezoelectric properties of the cochlear outer hair cell wall,'' J. Acoust. Soc. Am. **97**, 3006–3011.

White, N. R., Prasad, M., Barfield, R. J., and Nyby, J. G. (**1998**). ''40- and 70-kHz vocalizations of mice (Mus musculus) during copulation,'' Physiol. Behav. **63**, 467–473.

Zhao, H. B., and Santos-Sacchi, J. (**1999**). ''Auditory collusion and a coupled couple of outer hair cells,'' Nature (London) **399**, 359–362.

A time-selective technique for free-field reciprocity calibration of condenser microphones

Salvador Barrera-Figueroa,[a] Knud Rasmussen,[b] and Finn Jacobsen[c]
Technical University of Denmark, Ørsted-DTU, Acoustic Technology, Ørsted Plads, Building 352, DK-2800 Lyngby, Denmark

(Received 12 September 2002; revised 6 June 2003; accepted 8 July 2003)

In normal practice, microphones are calibrated in a closed coupler where the sound pressure is uniformly distributed over the diaphragm. Alternatively, microphones can be placed in a free field, although in that case the distribution of sound pressure over the diaphragm will change as a result of the diffraction of the body of the microphone, and thus, its sensitivity will change. In the two cases, a technique based on the reciprocity theorem can be applied for obtaining the absolute sensitivity either under uniform pressure or free-field conditions. In this paper, signal-processing techniques are considered that improve the accuracy of the free-field calibration method. In particular, a fast Fourier transform (FFT)-based time-selective technique for removing undesired reflections from the walls of the measurement chamber has been developed and applied to the electric transfer impedance function between two microphones. The acoustic centers of the microphones have been determined from the ''cleaned'' transfer impedance values. Then, the complex free-field sensitivities of the microphones have been calculated. The resulting complex sensitivities and acoustic centers have proved to be in good agreement with previously published data, and this confirms the reliability of the time-selective technique, even in nonanechoic environments. Furthermore, the obtained results give a new reference for further comparisons, because they cover a frequency range with an accuracy that has not been obtained by previously published data. [DOI: 10.1121/1.1604123]

PACS numbers: 43.38.Kb, 43.20.Ye, 43.60.Qv [DKW]

I. INTRODUCTION

The absolute determination of the free-field sensitivity of a condenser microphone is carried out making use of the reciprocity technique. The free-field reciprocity calibration of microphones was first discussed by MacLean.[1] Later, Wathen-Dunn[2] described in more detail some theoretical aspects of the calibration, such as the acoustic center of the microphones, and suggested that the coupling of the microphones may occur. Rudnick and Stein[3] presented an experimental study of the free-field calibration based on the works of MacLean and DiMattia.[4] Further experimental attempts of free-field calibrations were described by Terry[5] and Niemoller,[6] who introduced the possibility of carrying out calibrations in the time domain. In the 1990's, some national metrology institutes described the experimental apparatus used for free-field reciprocity calibrations, and presented some results.[7–11] Finally, the calibration method has been standardized.[12]

In the implementation of the free-field calibration setup a number of problems occur. The first is a very poor signal-to-noise ratio that can be negative at low frequencies when measuring the electrical transfer impedance between the two microphones. This can be partially solved by applying some signal conditioning techniques and by measuring the electrical transfer impedance at short distances between the microphones. However, it is worth mentioning that when the distance between the microphones is sufficiently short, a standing wave could be formed between the diaphragms of the microphones.

Another problem present in the free-field calibration is the nonideal performance of the anechoic environments developed to realize the free field under controlled environmental conditions. The problem is due to reflections from the walls of the anechoic chamber. These reflections and the standing wave between the microphones, though small, can have a significant influence on the measured electrical transfer impedance function between the microphones, and thus on the free-field sensitivity of the microphones. If a low uncertainty on the determined value of the sensitivity is to be achieved, the effect of such reflections must be reduced or eliminated from the electrical transfer impedance between the microphones. This can be determined in the time domain by means of a gating or time-selective technique either in real time or by postprocessing the data obtained in the frequency domain. Earlier work that includes time-selective techniques on the free-field calibration of condenser microphones has been done by Lambert and Durocher,[13] and Blem,[14] who explored the possibility of removing unwanted reflections by applying a time-selective window using an FFT-based procedure. Vorländer[15] also proposed the application of a time-selective window to the impulse response, which was obtained by the use of broadband signals. In this paper, an FFT-based time-selective technique is analyzed.

The proposed procedure consists of three major stages: (a) a treatment of the frequency response; (b) the determina-

[a]Electronic mail: sbf@oersted.dtu.dk
[b]Electronic mail: kr@oersted.dtu.dk
[c]Electronic mail: fja@oersted.dtu.dk

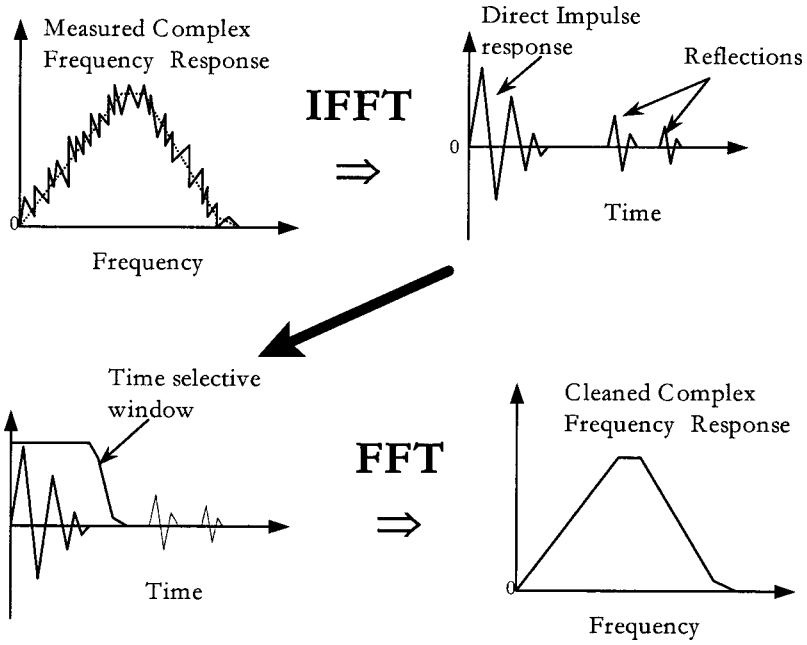


FIG. 1. Schematic of the procedure for removal of reflections of a complex frequency response function.

tion of the impulse response and the application of a time-selective window; and (c) the calculation of the cleaned frequency response.

It is possible to measure the electrical transfer impedance at equidistantly placed frequencies in a finite frequency range. If any Fourier transformation is to be applied, the missing portions of the frequency response must be completed by an adequate procedure. Then, the inverse Fourier transform is applied to the completed frequency response in order to obtain the impulse response. It contains information about the direct wave between the microphones as well as the reflections and standing waves. These undesired effects can be removed by means of a time-selective window.

Finally, the cleaned electrical transfer impedance can be obtained by applying the Fourier transform. This result can be used for determining the acoustic centers of the microphones and the free-field sensitivity of the condenser microphone. A schematic representation of the procedure is shown in Fig. 1.

The overall effect of the application of the cleaning procedure over the accuracy of the microphone sensitivity is studied by making use of a simulated transfer impedance between the two microphones. These results are used for validating the application of the time-selective technique by comparing them with complementary experimental and numerical results.

II. THE ELECTRICAL TRANSFER IMPEDANCE

Consider two microphones located at an acoustical distance d_{12} from each other in a free field. The electrical transfer impedance between the microphones at the frequency f is defined in Ref. 12 as the ratio of the open-circuit voltage on the electrical terminals of the receiver microphone, u_2, to the electrical current through the electrical terminals of the second microphone acting as sound source, i_1

$$Z_{e,12}=\frac{u_2}{i_1}=j\,\frac{\rho f}{2d_{12}}M_{f,1}M_{f,2}\exp(-j\gamma d_{12}), \tag{1}$$

where $M_{f,1}$ and $M_{f,2}$ are the free-field sensitivities of the microphones, ρ is the density of the medium, and γ is the complex propagation coefficient, which includes the air absorption. The electrical transfer impedance is the basis of the reciprocity technique as it contains information about the free-field sensitivities of the two microphones, thus providing means for determining their sensitivities. Figure 2 shows a diagram of the microphones in the free field.

Information regarding the shape of the corresponding impulse response can be extracted from the factors in Eq. (1). It is possible to divide the electrical transfer impedance into three different factors. The first is the frequency multiplied by a constant, the second is the product of sensitivities, and the third is the complex exponential of the product of acoustical distance and the complex propagation coefficient. The

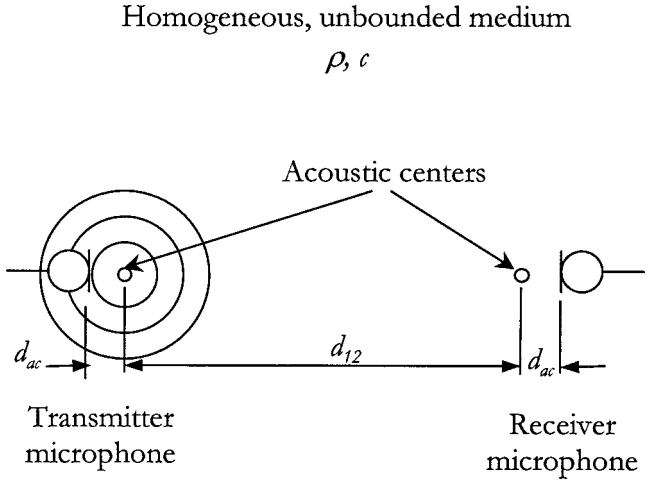


FIG. 2. Schematic of two microphones in a free field. One of the microphones is acting as a sound source (transmitter) and the other as a receiver.

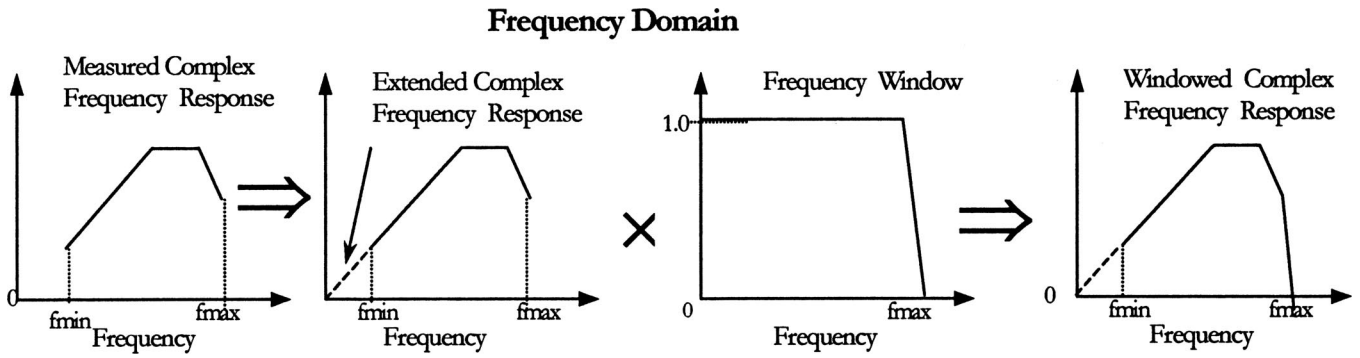


FIG. 3. Schematic of the procedure for the treatment of an incomplete frequency response prior to the Fourier analysis.

shape of the impulse response of the electrical transfer impedance is the convolution of the impulse responses of each factor.

The first term may be considered as a differentiation that emphasizes high frequencies; this is a consequence of the fact that the transmitter microphone is acting as a monopole. The second term can be considered as a multiplication of two systems, each of a single degree of freedom. The last term can be interpreted as a time delay. This implies that the variations observed in the impulse response should be mostly associated with the variations in the parameters of the two single-degree-of-freedom systems associated with the free-field sensitivities. This indicates that the parameters of the time-selective procedure should be selected taking such information as a basis.

A quantity used for determining the time it takes the impulse response of a system of a single degree of freedom to decay to an arbitrary level with respect to its maximum amplitude is the time constant, τ, which is given by

$$\tau = \frac{2m_a}{R_a}, \tag{2}$$

where m_a is the acoustical mass and R_a is the acoustical resistance from the system of a single degree of freedom.

III. REMOVAL OF REFLECTIONS

The reflections from the walls of the anechoic chamber and standing waves between the microphones can be eliminated from the electrical transfer impedance, $Z_{e,12}(f)$, using the following procedure. The impulse response of the electrical transfer impedance, $z_{e,12}(t)$, can be calculated by applying the inverse Fourier transform (IFT). In the time domain, the reflections appear as scaled and perhaps distorted replicas delayed from the main impulse response. Thus a time-selective window, $w(t)$, may be applied to remove them. Then, the ''cleaned'' electrical transfer impedance, $Z_{e,12,C}(f)$, can be calculated by applying the Fourier transform to the windowed impulse response, $z_{e,12,w}(t)$. The procedure graphically shown in Fig. 1 can be formalized as

$$Z_{e,12}(f) \xrightarrow{\text{IDFT}} z_{e,12}(t) \xrightarrow{w(t)} z_{e,12,w}(t) \xrightarrow{\text{DFT}} Z_{e,12,C}(f). \tag{3}$$

According to the definition of the IFT of a complex frequency response, $X(f)$

$$x(t) = \int_{-\infty}^{\infty} X(f)\exp(j2\pi ft)df, \tag{4}$$

the frequency response must be known at all frequencies. However, in practice, it is only possible to measure the electrical transfer impedance in a limited frequency range. There is a low-frequency limit, f_0, below which it is impossible to measure the electrical transfer impedance. This is caused by the microphone's thermal noise.[16] When combined with the microphone's very low radiation capabilities, this results in an extremely poor signal-to-noise ratio at low frequencies. On the high-frequency side, the limitations come from the capacity of the data acquisition software, which makes it possible to measure only up to a frequency $f_{\max}$ where the electrical transfer impedance may not have decayed sufficiently.

If any Fourier transform-based postprocessing is to be performed on the electrical transfer impedance, it has to be defined in the whole frequency interval, formally from $-\infty$ to $+\infty$ as in Eq. (4), or from 0 to $+\infty$ for a one-sided frequency response. Then, a treatment of the missing portions of the frequency response must be carried out.

A. Treatment of the frequency response

The frequency response must be completed in the whole frequency domain before the inverse Fourier transform is applied. This means that the low-frequency portion $[0,f_0]$ and the high-frequency portion $[f_{\max},+\infty]$ must be generated in such a way that follows the physical behavior of the function without modifying the measured electrical transfer impedance function. The procedure is shown schematically in Fig. 3.

1. Low-frequency artificial extension of the electrical transfer impedance

The missing low-frequency values can be completed by patching ideal values to the measured of the electrical transfer impedance. This solution is based on the fact that at low frequencies, diffraction and radiation impedance effects are almost negligible, and the free-field sensitivity becomes constant and almost equal to the pressure sensitivity. Thus, a lumped-parameter model can be used for generating the ideal electrical transfer impedance.[17] The lumped parameters can be determined from pressure sensitivity data. The resulting patched electrical transfer impedance, $Z_{e,12,E}(f)$, is defined as

$$Z_{e,12,E}(f)=\begin{cases} Z_{e,12,i}(f) & 0\leqslant f\leqslant f_0 \\ Z_{e,12,m}(f) & f_0\leqslant f\leqslant f_{\max}, \end{cases} \tag{5}$$

where $Z_{e,12,i}$ is the ideal electrical transfer impedance calculated from the lumped-parameter model of the microphone and $Z_{e,12,m}(f)$ is the measured electrical transfer impedance.

One of the possible shortcomings of this procedure is that the effect of the reflections is larger at low frequencies. As the procedure uses an ideal, reflectionless frequency response for filling the missing part, a realistic representation of the disturbances may not be obtained, though ultimately the reflections are to be removed. Furthermore, the patching of ideal data to measured data may introduce an additional problem. The value of the electrical transfer impedance at the lowest measured frequency, f_0, may correspond either to a maximum or a minimum of the standing wave pattern. This will introduce a discontinuity on the slope of the extended electrical transfer impedance function that may be reflected onto the impulse response.

2. Low-pass filtering of the electrical transfer impedance

The sensitivity of a microphone tends to decay at high frequencies as a consequence of the mechanical behavior of the diaphragm. This is explained by the fact that above the resonance frequency of the diaphragm, its movement is controlled by its mass, and therefore the sensitivity approaches zero asymptotically. This behavior is also reflected in the electrical transfer impedance. It suggests the possibility of accelerating artificially this decay in the measured frequency range, provided that the upper frequency is well above the resonance frequency of the microphones. This can be done with a low-pass filter. The low-pass filtered electrical transfer impedance, $Z_{e,12,L}(f)$ is defined by

$$Z_{e,12,L}(f)=Z_{e,12,E}(f)\cdot L(f), \tag{6}$$

where $L(f)$ is the low-pass filter. This filter should not alter the modulus of the electrical transfer impedance in the frequency range of concern, and therefore, ideally, it should have a value of unity in the interval $[0,f_f]$ and zero in the rest of the whole frequency range. However, an ideal filter with these characteristics will introduce a noncausality problem on the impulse response; therefore, a causal filter should be applied instead. Such a filter, a linear-phase FIR filter for instance, can be developed by means of the techniques described in several digital signal-processing textbooks, for example in Ref. 18.

B. Time-selective window

Once the missing portions of the electrical transfer impedance have been completed, the inverse Fourier transform is applied and the impulse response, $z_{e,12}(t)$, is obtained. The impulse response contains information about the direct wave between the two microphones, the standing wave between the microphones and the reflections from the walls of the anechoic room. In order to remove the last two effects, a time-selective window, $w(t)$, can be applied. This time-selective window must not modify the direct wave between the microphones, and must completely eliminate the reflections, which can be considered as attenuated, delayed replicas of the direct wave. Therefore, the time-selective window should have a value of unity in the time interval where the impulse response contains the direct wave, and zero everywhere else. The first choice may be a rectangular window. However, as this window has high sidelobes, it is expected that it will introduce some ripple in the frequency response caused by cutting abruptly the impulse response where it has not decayed completely to zero. This effect can be reduced by applying a smoothing function on the extremes of the time window. A window with smooth extremes can be generated by convolving the rectangular window with a smoothing function. Although the extremes of the window can be considerably smoothed by the convolution, the height of the first sidelobe is approximately the same as that of the rectangular window. This is a natural consequence of the generation and form of the window itself. One of the convolved windows is the Tukey window[19]

$$w_{\text{Tukey}}(n)=\begin{cases} 1 & 0\leqslant |n| \leqslant (1+\alpha)N/2 \\ \dfrac{1}{2}\left\{1+\cos\left[\pi\dfrac{n-(1+\alpha)N/2}{(1-\alpha)N/2}\right]\right\} & \\ \quad (1+\alpha)N/2\leqslant |n| \leqslant N & \\ 0 \quad \text{elsewhere}, & \end{cases} \tag{7}$$

where α is the smoothing portion of the window, and N is the length of the window. The convolution procedure can also be applied for other smoothing functions such as those related to Kaiser or Blackman windows.[19]

IV. CALCULATION OF THE SENSITIVITY AND OTHER PARAMETERS

Once the undesired reflections on the electrical transfer impedance have been removed by the procedure described above, it is possible to use this cleaned function for the calculation of the sensitivity of the microphones and other associated parameters.

The free-field sensitivity of a condenser microphone can be obtained from Eq. (1). This equation contains the product of the free-field sensitivities of the two microphones. It provides an equation with two unknowns. If a third microphone is coupled successively with the other two microphones, a set of three equations with three unknowns is obtained. Then, the sensitivity of each microphone can be obtained in absolute terms. Solving the system of equations, the sensitivity of the first microphone is given in Ref. 12 as

$$M_{f,1}=\frac{2}{\rho f}\frac{d_{12}d_{13}}{d_{23}}\frac{Z_{e,12}Z_{e,13}}{Z_{e,23}}\exp[\gamma(d_{12}+d_{13}-d_{23})]. \tag{8}$$

Similar expressions can be obtained for the other two microphones. This equation is used for calculating the free-field sensitivity after cleaning the electrical transfer impedance functions. Also, it can be proved (see Ref. 12) that the free-field sensitivity can be expressed as a function of the product of the pressure sensitivity (where the diaphragm of the microphone is subjected to a uniform pressure), a diffraction factor $S(f,\theta)$, and the ratio of the impedance of the micro-

phone, Z_a, to the radiation impedance $Z_{a,r}$, as follows:

$$M_f = M_p \frac{Z_a}{Z_a + Z_{a,r}} S(f,\theta). \tag{9}$$

The latter two factors can be considered as a single one, and renamed as the free-field factor, which is the ratio of the free-field sensitivity to the pressure sensitivity. This quantity is of great practical utility, and it is widely used in field measurements in its logarithmic form and known as the free-field correction. It provides a parameter that is easier to compare with experimental and theoretical results obtained elsewhere. Also, as the most significant effect shown in Eq. (9) is the diffraction factor, it is expected that the free-field correction should be the nearly the same for all microphones with the same geometrical configuration and acoustical impedance.

It can be seen that in Eq. (8) there is a parameter with a significant influence on the final free-field sensitivity. This parameter is the true acoustic distance between the microphones. The true acoustic distance depends on the acoustic centers of the two microphones as defined in Ref. 12.

The experimental determination of the acoustic center can be carried out using simple linear regression techniques if measurements are made at several distances. This procedure can be used when there is confidence in the characteristics of the air in the environment where the measurements are made.

V. COMPUTER MODELING OF THE ELECTRICAL TRANSFER IMPEDANCE

A computer model based on Eq. (1) has been developed using MATLAB. The product of sensitivities in the electrical transfer impedance is generated by making use of the lumped-parameter model of the microphones and a typical free-field correction.[20] The reflections from the imperfect anechoic chamber and the standing wave between the microphones are introduced using the concept of image sources. It should be mentioned that Delany and Bazley[21] have pointed out that the total field at any point inside the enclosure should be calculated by summing the six waves reflected from the walls and the direct wave from the source with due regard of the relative phase of these waves, and that their relative phase may change as a function of the impedance of the walls. This should give a better agreement between the calculated and measured interference patterns. On the other hand, the walls of the anechoic chamber are formed by a number of wedges that will certainly modify the shape of the reflected wave in an unpredictable manner. However, in this case it is not very important to reproduce the exact shape of the reflections, but to indicate the approximate time instant when they occur. This can be achieved with the use of the image source theory. The cleaning procedure is applied to this simulated electrical transfer impedance in order to evaluate the effect of the procedure as well as of the signal-processing parameters used to generate it. The selection of such parameters is described below.

The first aspect to be analyzed is that of the duration of the direct wave between the microphones. This can be estimated by making use of the lumped-parameter model of the microphones and the time constant defined in Eq. (2).

TABLE I. Values of the lumped parameters for condenser microphones LS1P.

	LS1 (B&K 4160)
Acoustic mass, m_a	345 kg m^{-4}
Acoustic compliance, C_a	1.19×10^{-12} m^3 Pa^{-1}
Acoustic resistance, R_a	20×10^{-6} Pa s m^{-3}

The microphones to be analyzed are laboratory standard microphones: 1-in. laboratory standard microphones, LS1P. The above type corresponds to the microphone Brüel and Kjær model 4160. The results from this study can be extended to other types of microphones as for instance the 1/2-in. laboratory standard, LS2P. Typical values given by the manufacturer for the lumped-parameter model of the pressure sensitivity are shown in Table I.

The impulse response will have a peak at the instant corresponding approximately to the time it takes for the sound wave to travel from the transmitter microphone to the receiver. After this peak it will decay exponentially according to the time constant defined in Eq. (2). This constant provides a measure of how rapidly the amplitude of an impulse response of a system of a single degree of freedom decays following an exponential behavior, $e^{-(1/\tau)t}$, where τ is the time constant, and t is the time. This can be used to predict the time when a given decay is expected to occur. Thus, for the parameters given in Table I, the time constant is 0.0345 ms. This value indicates a fast decay of the impulse response. Note that Eq. (1) contains a product of two systems of a single degree of freedom. Then, the resulting convolution has a time constant that is two times the time constant of one system. Thus, using the calculated time constant, the direct wave should have decayed to one thousandth of its maximum amplitude about 0.5 ms after the peak has been reached. However, it is worth mentioning that it is expected that the actual values of the lumped-parameter model may change under free-field conditions, due to the presence of a radiation impedance and the diffraction factor. Such changes may modify the resonance frequency and the quality factor of the microphone.

With the above result it is possible to determine whether the reflections from the walls and the expected standing wave between the microphones will be far enough from the direct wave to avoid truncating it when applying the time-selective window. Consider that the microphones are located at a distance of 30 cm from each other in the middle of an anechoic room having free-space dimensions of $120\times80\times175$ cm. At these conditions, the direct wave between the microphones will have its peak at about 0.87 ms, and the amplitude of the impulse response should decay to one thousandth of its maximum at 1.37 ms. On the other hand, the closest disturbance to the direct wave is the standing wave between the microphones. The time instant where it should appear in the impulse response is about 2.6 ms. The rest of the reflections will come at a later time: first from the lateral walls and then from the ceiling and the floor of the chamber. These last two reflections come at a time of about 5.1 ms, but

as discussed above the actual reflection point on the walls is unknown, and therefore the times given above may change in the actual measurements.

In order to obtain a realistic impulse response, the frequency response should be sampled using parameters that are in full accordance with the physical characteristics of the microphones. The resonance frequency and the quality factor of the microphones, Q, give information about the lowest higher-limit frequency, $f_{\max}$, that must be measured or simulated in order to determine a realistic impulse response. For LS1P microphones, the resonance frequency is about 8.5 kHz, and the quality factor is nearly unity. This implies that the system has a broad resonance. And, at a frequency 1.5 times the resonance frequency, the average power has dropped 50% or 3 dB the from the value at the resonance frequency; if a realistic decay of the impulse response is to be achieved, the maximum frequency should not be less than the above limit. It must be also noticed that the Eq. (1) contains the frequency as a factor, which corresponds to a differentiation that clearly emphasizes high frequencies. As it is desirable to have a frequency response whose high-frequency values should have decayed sufficiently, a proper minimum value for $f_{\max}$ should be estimated.

As the maximum frequency defines the size of the time step in the corresponding impulse response, the number of sampled frequency points should be large enough (a) to describe properly the irregularities caused by the standing wave and reflections from the walls, and (b) to include all the primary reflections, i.e., to give a time record that includes such reflections.

The first reflection appears with a frequency periodicity of about 380 Hz, and the latest with a period of about 200 Hz. This implies that in order to sample the disturbances appropriately, the size of the frequency step should not be larger than 100 Hz, and less if possible. This figure may change slightly if the distance between the microphones is increased, as it is in practical situations where it can take values between 20 and 40 cm. This is a minor variation, however, because the time where the reflections appear does not change in a significant manner when the distance between the microphones is in the given interval.

The number of frequency samples is given by dividing $f_{\max}$ by the frequency step. The number of frequency points as well as the sampling frequency can be increased in order to obtain a sampling of the shape of the disturbances. These calculated time parameters have been used as a basis for a number of simulations of the cleaning procedure applied to a simulated electrical transfer impedance that contains reflections from the walls.

The first parameter to be evaluated is the upper frequency, $f_{\max}$. The objective is to provide a suitable value of $f_{\max}$ so as to generate a realistic impulse response that contains all the relevant information required for a proper separation of the direct wave between the microphones. For this purpose, an electrical transfer impedance function between two LS1P microphones located 30 cm from each other has been generated at four different frequency ranges with maximum frequencies, $f_{\max}$, of $1.5\cdot f_{\text{res}}$, $2\cdot f_{\text{res}}$, $3\cdot f_{\text{res}}$, and $4\cdot f_{\text{res}}$. In all four cases the frequency interval has been divided into 1024 points. A low-pass frequency window has been applied onto the simulated electrical transfer impedance in order to make it converge to zero at the high-frequency limit. Figure 4 shows a graphic of the electrical transfer impedance function for the last case, $4\cdot f_{\text{res}}$. Figure 5 shows the Hilbert envelope of the corresponding impulse responses of the four cases.

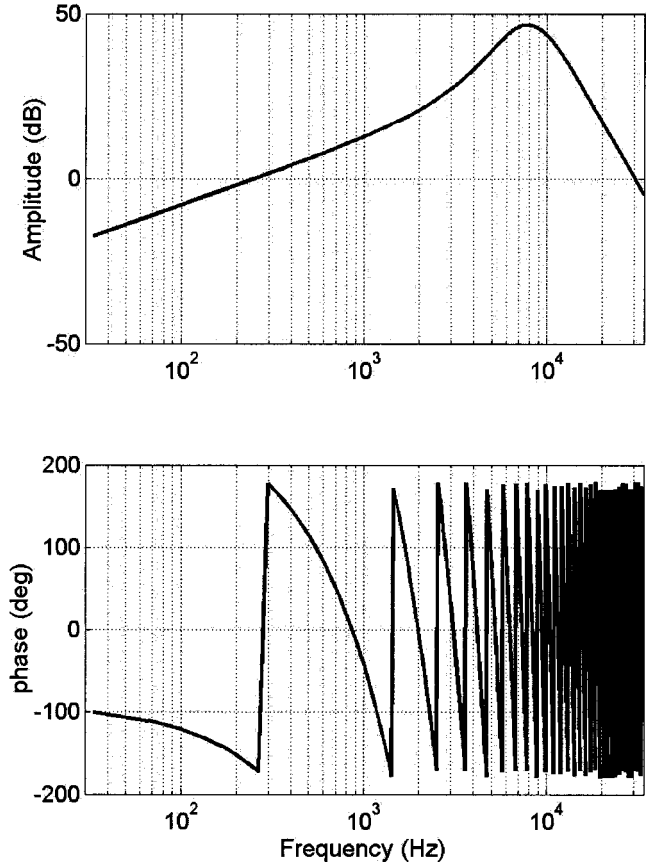


FIG. 4. Simulated electrical transfer impedance function between two LS1P microphones located 30 cm from each other in a free field. The maximum frequency is $4\cdot f_{\text{res}}$.

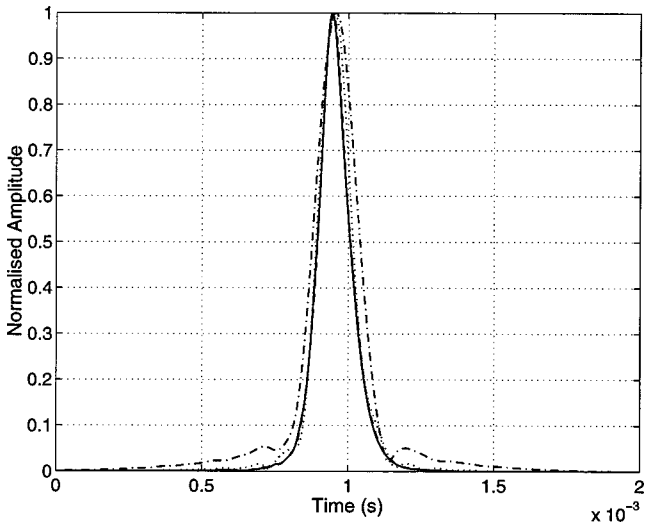


FIG. 5. Hilbert envelope of the impulse response from four simulated electrical transfer impedance functions with maximum frequencies: ——— $4\cdot f_{\text{res}}$; --- $3\cdot f_{\text{res}}$; ··· $2\cdot f_{\text{res}}$, and -·- $1.5\cdot f_{\text{res}}$.

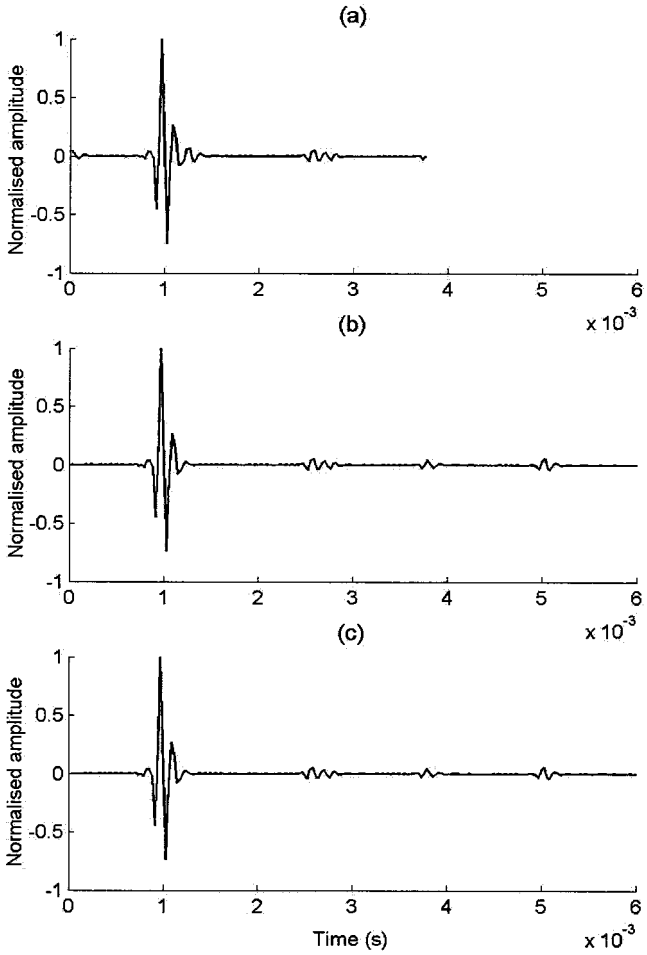


FIG. 6. Impulse response of the simulated electrical transfer impedance function for different sizes of the frequency step: (a) 264 Hz; (b) 132 Hz; and (c) 66 Hz.

As expected, the direct wave between the microphones rises to a maximum after about 0.9 ms. It can be seen that in the case where $f_{\max}=1.5\cdot f_{\text{res}}$, the impulse response rises and decays slower than under the other three conditions. This has an important consequence if the direct wave is to be isolated from the reflections. Additionally, a small time shift of a fraction of a millisecond can be observed between the peaks of the four cases; the difference tends to converge as the value of $f_{\max}$ increases. This is basically a consequence of the major high-frequency information contained in the frequency responses with a higher $f_{\max}$ and has no important consequence on the overall procedure.

The second parameter to be evaluated is the size of the frequency step. The intention is to find a suitable step size that will represent accurately the effect of the reflections in the impulse response. This is important in this case because the measurements of the electrical transfer impedance are made *frequency by frequency*, and thus the number of frequency steps should be limited. The simulated electrical transfer impedance corresponds to the case of two LS1P microphones located at a distance of 30 cm from each other. The upper frequency $f_{\max}$ is chosen to be two times the resonance frequency of the microphones. Artificial reflections have been added by introducing mirror sources. Three cases have been studied, dividing the frequency interval into (a) 64; (b) 128; and (c) 256 frequency steps that corresponds to frequency steps of 264, 132, and 66 Hz, respectively. The impulse responses corresponding to the three cases are shown in Fig. 6.

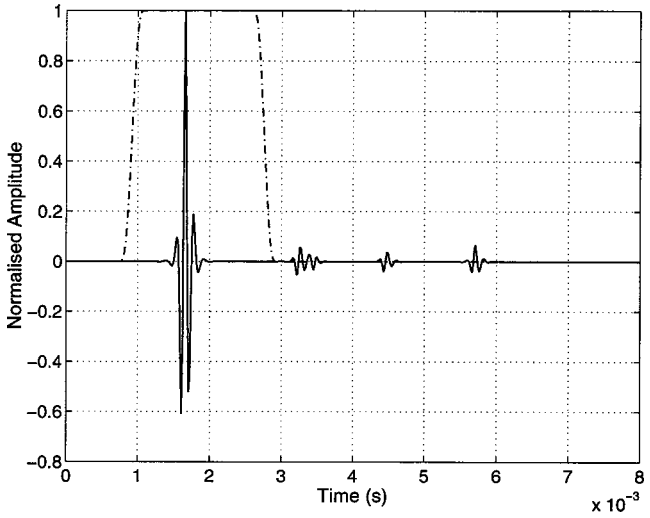


FIG. 7. Impulse response of the simulated electrical transfer impedance function and the time-selective window: ——— impulse response, and -·- time-selective window.

It can be seen that the step size of the frequency response is an important parameter. As expected, the large size of the frequency step in case (a) leads to a subsampling of the disturbances caused by the latest of the reflections. This can be seen in Fig. 6(a), where the corresponding impulse response is not long enough to include a portion of the reflections, and even at zero time a portion of the last reflection is present. This may be caused by an aliasing effect. Case (b) describes a situation that is on the limit of aliasing, but it contains all the reflections. Finally, case (c) is a frequency interval short enough to avoid any aliasing. From the above results it can be concluded that it is possible to carry out a simulation of the cleaning procedure of the electrical transfer impedance describing accurately its impulse response. It is intended to assess experimental measurements with the simulation results, and therefore sampling parameters derived from the above methodology that are compatible with those used in the experimental setup are to be used in the following.

The electrical transfer impedance between two LS1P microphones has been generated by using 993 frequency steps in the frequency range from 900 to 30 690 Hz, i.e., with a frequency step of 30 Hz. Artificial reflections have been introduced. Background noise has been introduced by adding a random variation to the simulated electrical transfer impedance of the same level to that observed in the measurement setup. The missing portion of the contaminated electrical transfer impedance is filled with values of a reflectionless electrical transfer impedance. Then, a low-pass filter is applied to the frequency response. Thereafter, the impulse response is obtained, and a Tukey window [Eq. (7)] is applied. These are shown in Fig. 7. The peak of the impulse response is expected to occur at 0.87 ms; however, the linear-phase characteristic of the FIR low-pass filter introduces the additional delay that is observed in Fig. 7.

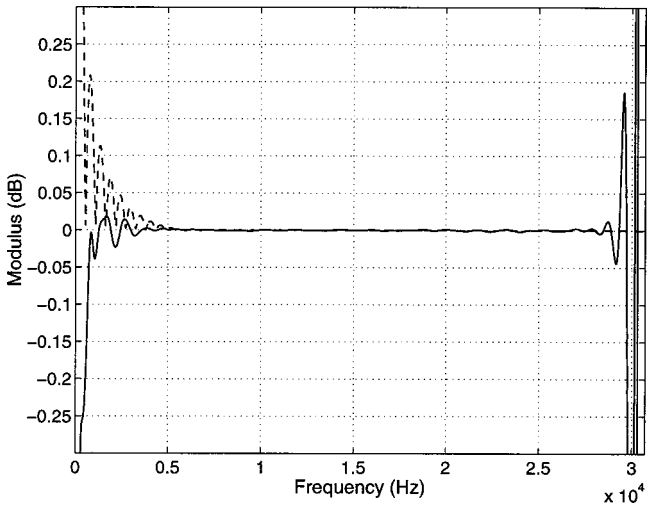


FIG. 8. Simulated difference between the cleaned and an ideal electrical transfer impedance function as a measure of the global effect of the cleaning procedure on the accuracy of the electrical transfer impedance. ——— residual; --- spectrum of the time-selective window.

The most important outcome of this simulation study is the effect caused by the application of the cleaning procedure described in Eq. (3). For this purpose, the cleaned electrical transfer impedance is compared with an ideal reflectionless electrical transfer impedance; the overall effect of the procedure is defined as the difference between the two quantities. This result is to be used for assessing the accuracy improvements caused by the removal of the reflections and any deviation introduced by the procedure. Figure 8 shows the difference between the cleaned electrical transfer impedance and the reflectionless electrical transfer impedance. It can be seen that major differences are present at the extremes of the residual function. These differences are apparently caused by the truncation of the impulse response at an instant when it has not decayed sufficiently. The ripple on the residual is explained by the fact that the time-selective window has a short duration compared with the total length of the impulse response. Thus, its Fourier transform has high and wide sidelobes; it can be seen at the residual function that the frequency of the distortions at the extremes coincides with that present in the Fourier transform of the time window, whose modulus is also shown in Fig. 8.

Another effect that can be observed is that at the patching frequency, f_0, there is a discontinuity of amplitude and slope between the ideal and the contaminated electrical transfer impedance functions. This is caused by the fact that the ideal value of the frequency response at that frequency may coincide with a maximum or a minimum of the contaminated frequency response. Though this may introduce an additional harmonic distortion, this may not be significant because the patching level is normally 40 dB lower than the maximum value of the frequency response.

All in all, it can be concluded that the most significant deviations introduced by the cleaning procedure are caused by the time-selective window. The frequency range where the time window introduces the maximum levels of distortion can be determined from its frequency spectrum by evaluating the width and height of the main and secondary lobes. Thus, the accuracy remains unchanged in the frequency range where the distortion caused by the time-selective window can be considered as negligible. For metrological purposes, this can be set to be compared with a given uncertainty level.

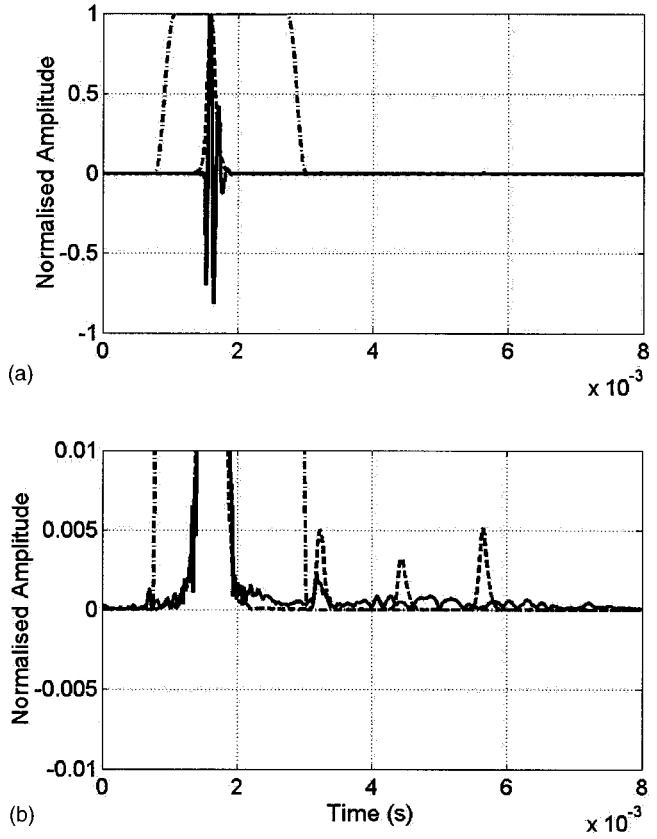


FIG. 9. Impulse response obtained from experimental data. (a) is the normalized amplitude and the time window: ——— Impulse response, --- Hilbert envelope of the ideal impulse response; and -·- time window. (b) ——— Hilbert envelope of the measured response, --- Hilbert envelope of the ideal impulse response; and -·- time window.

VI. EXPERIMENTAL RESULTS

In this section, the cleaning procedures have been applied on experimental data from LS1P microphones. The experimental setup consists of three LS1P standard microphones, B&K 4160, placed in a small anechoic room with free space dimensions of $120\times80\times175$ cm. The measurement of the frequency response is made using the so-called steady-state response mode of the analyzer, B&K 2012. The analyzer is connected to a reciprocity apparatus that measures the voltage on the terminals of the receiver microphone and the current through the terminals of the transmitter microphone.

Figure 9 shows the impulse response of an electrical transfer impedance between two microphones located at a distance of 28 cm from each other. It can be seen that the measured impulse response and the ideal impulse response have very similar characteristics in terms of duration and decay time, thus validating the simulation results. Also, the position of the first reflection which is caused by the standing wave between the microphones coincides in the two cases.

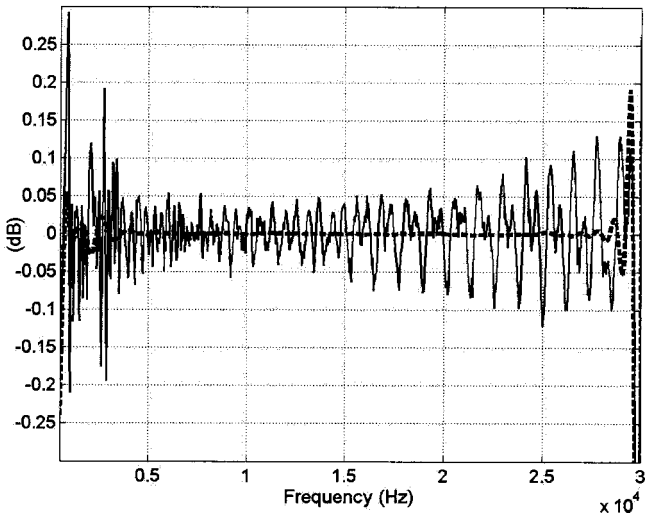


FIG. 10. Difference between the cleaned and the raw experimental electrical transfer impedance function compared with the difference between the cleaned and ideal electrical transfer impedance functions obtained from the simulations. ——— Difference between cleaned and raw, --- difference between ideal and cleaned.

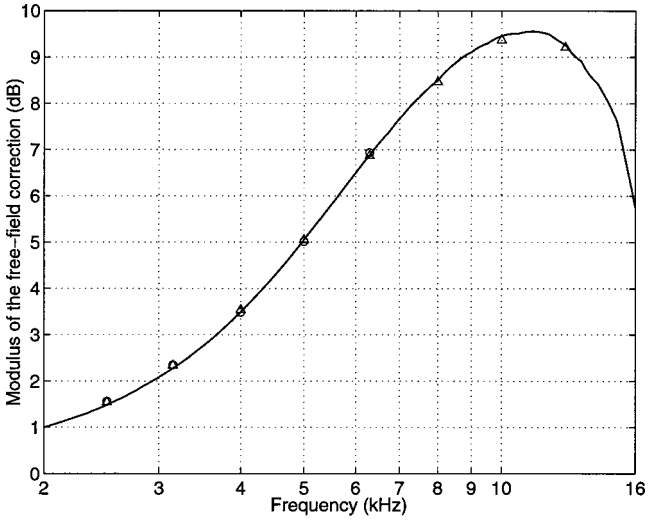


FIG. 12. Experimental free-field correction for LS1P condenser microphones. ——— Average value for three microphones, ○ Juhl (Ref. 22), and △ IEC655 standard (Ref. 20).

However, the reflections from the walls appear to come later in the measured data; this may be a consequence of the selection of the reflection point at the wall, which in the simulation is considered to be the tip of the wedges. A more realistic simulation should consider the complicated reflections coming from the wedges.

In order to assess the effectiveness of the technique for removing the reflections, it is necessary to compare the size of the removed disturbances with the size of the distortion introduced by the cleaning procedure (shown in Fig. 8). In Fig. 10 the difference between the cleaned and the measured electrical transfer impedance function for the same case is shown. In Fig. 11 the result of the acoustic centers determined from the cleaned and raw measurements are shown. The acoustic centers were obtained using a linear regression procedure on a set of measurements of the electrical transfer impedance at different distances. These results are compared with data provided by the Refs. 11, 12, and 22. It is apparent that the uncleaned or raw measurements give large variations of the acoustic centers. These deviations are not inherent to the acoustic centers but a consequence of the interference patterns of the direct and reflected waves. The approximation obtained using the cleaned electrical transfer impedances appears to follow the values given for comparison better, although it diverges at frequencies above 10 kHz. These deviations could be caused by the fact that the acoustic centers are strongly dependent on the velocity distribution of the diaphragm, and around and above the resonance frequency very little can be assumed about the behavior of this quantity.

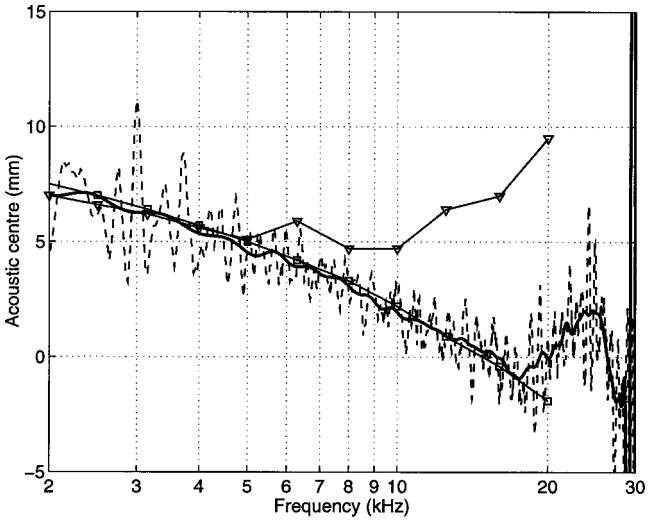


FIG. 11. Position of the measured acoustic centers of three LS1P condenser microphones: ——— Average of the cleaned for three microphones, --- raw, -□- IEC standard (Ref. 12), and -▽- Rasmussen (Ref. 11).

Another fact that is worth mentioning is the difference between the cleaned acoustic centers and the values given in Ref. 11. The present measurements and those of Ref. 11 were carried out using the same facilities. At the time when this report was written, it was suspected that the difference might be due to an anomalous air absorption inside the anechoic chamber. In the course of the investigation, a number of measurements were made while the door of the small anechoic chamber was open. The results obtained were similar to those given in the standard.[12] After a series of additional tests, a ventilation system was introduced in the chamber. The results presented here correspond to the new condition, which proves that the air inside the chamber was causing the divergence. Thus, application of the cleaning technique has helped to solve some discrepancies observed between the theory and experimental measurements. Additionally, it shows the importance of verifying that the physical properties of the air inside the anechoic environment where free-field measurements are to be made are those of standard air.

Figure 12 shows the average free-field correction of the three microphones calibrated and values given in Ref. 22 and the standard.[20] The results shown in Fig. 12 are in good agreement with the previously published data. Unfortunately, because there is not much information available for the high-

frequency range (above 10 kHz), the obtained results cannot be compared with another reference.

VII. CONCLUSIONS

A time-selective procedure has been developed and tested using computer simulations and experimental data. The procedure has proved to remove the imperfections on the electrical transfer impedance caused by reflections effectively. It makes it possible to reduce the uncertainty of the free-field sensitivity.

The acoustic centers of the microphones have been determined experimentally with improved accuracy over an extended frequency range that has not been covered before. The determined values show good agreement with previously published values at frequencies below and around the resonance frequency. The high-frequency values provide a new reference for further analysis. Additionally, the obtained values have helped to solve some discrepancies observed in previously published works.

The free-field sensitivity was calculated from the cleaned electrical transfer impedances using the experimental acoustic centers. The low-frequency values of free-field sensitivity and, thus, the free-field correction are also in agreement with previously published data. The accuracy of the calculated sensitivity has been improved significantly in an extended frequency range not covered in previous work. The time-selective technique makes it possible to remove reflection artifacts from the data, thereby effectively obtaining free-field calibration conditions.

ACKNOWLEDGMENTS

This research project is carried out with the support of the Consejo Nacional de Ciencia y Tecnología (CONACYT) and the Centro Nacional de Metrología (CENAM) of México.

[1] W. R. MacLean, ''Absolute measurement of sound without a primary standard,'' J. Acoust. Soc. Am. **12**, 140–146 (1940).
[2] W. Wathen-Dunn, ''On the reciprocity free-field calibration of microphones,'' J. Acoust. Soc. Am. **21**(5), 542–546 (1949).
[3] I. Rudnick and M. N. Stein, ''Reciprocity free field calibration of microphones to 100 Kc in air,'' J. Acoust. Soc. Am. **20**(6), 818–825 (1948).
[4] A. L. DiMattia and F. M. Wiener, ''On the absolute pressure calibration of condenser microphones by the reciprocity method,'' J. Acoust. Soc. Am. **18**(2), 341–344 (1946).
[5] R. L. Terry and R. B. Watson, ''Pulse technique for the reciprocity calibration of microphones,'' J. Acoust. Soc. Am. **23**(6), 684–685 (1951).
[6] A. F. Niemoeller, ''Reciprocity calibration of condenser microphones in the time domain,'' J. Acoust. Soc. Am. **33**(12), 1712–1719 (1961).
[7] D. L. H. Gibbings and A. V. Gibson, ''Free-field reciprocity calibration of capacitor microphones at frequencies from 19.95 kHz to 316.2 Hz,'' Metrologia **20**, 85–94 (1984).
[8] E. W. Burnett and V. Nedzelnitsky, ''Free-field reciprocity calibration of microphones,'' J. Res. Natl. Bur. Stand. **92**(2), 129–151 (1987).
[9] J.-N. Durocher, ''Etalonnage des microphones à condensateur en champ libre,'' J. Acoust. **2**, 431–436 (1989).
[10] R. Barham, ''Free-field reciprocity calibration of laboratory standard microphones,'' Ph.D thesis, ISVR, University of Southampton, 1995.
[11] K. Rasmussen and E. Sanderman Olsen, ''Intercomparison on free-field calibration of microphones,'' The Acoustics Laboratory, Technical University of Denmark, Report PL-07 (1993).
[12] IEC 61094-3 (1995-11), ''Measurement microphone Part 3: Primary method for free-field calibration of laboratory standard microphones by the reciprocity technique'' (1995).
[13] J. M. Lambert and J. N. Durocher, ''Analyse des perturbations acoustiques lors de l'etalonnage en champ libre des microphones étalons à condesateurs dits d'un pouce par la technique de la réciprocité'' (in French), Laboratoire National D'Essais (1989).
[14] J. S. Blem, ''Tidsselektiv fritfeltskalibrering af kondensatormikrofoner'' (in Danish), Master thesis, The Acoustics Laboratory, Technical University of Denmark (1992).
[15] M. Vorländer and H. Bietz, ''Novel broad band reciprocity technique for simultaneous free-field and diffuse-field microphone calibration,'' Acustica **80**, 365–377 (1994).
[16] V. Tarnow, ''Thermal noise in microphones and preamplifiers,'' Brüel & Kjær Technical Review, 3–14 (1972) (3).
[17] K. Rasmussen, ''The static pressure and temperature coefficients of laboratory standard microphones,'' Metrologia **36**, 265–273 (1999).
[18] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing. Principles, Algorithms and Applications*, 3rd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1996).
[19] F. J. Harris, ''On the use of windows for harmonic analysis with the discrete Fourier Transform,'' Proc. IEEE **66**(1), 51–83 (1978).
[20] IEC International Standard 655, ''Values for the difference between free-field and pressure sensitivity levels for one-inch standard condenser microphones'' (1979).
[21] M. E. Delany and E. N. Bazley, ''A note on the sound field due to a point source inside and absorbent-lined enclosure,'' J. Sound Vib. **14**(2), 151–157 (1971).
[22] P. Juhl, ''A numerical investigation of standard condenser microphones,'' J. Sound Vib. **177**(4), 433–446 (1994).

Drum silencer with shallow cavity filled with helium

Y. S. Choy and Lixi Huang[a)]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

(Received 12 January 2003; revised 6 June 2003; accepted 30 June 2003)

The motivation of this study is twofold: (a) to produce a flow-through silencer with zero pressure loss for pressure-critical applications, and (b) to tackle low frequency noise with limited sideway space using cavities filled with helium. The work represents a further development of our recently conceived device of a drum-like silencer with conventional air cavity [Huang, J. Acoust. Soc. Am. **112**, 2014–2025 (2002); Choy and Huang, *ibid.* **112**, 2026–2035 (2002)]. Theoretical predictions are validated by experimental data. The new silencer consists of two highly tensioned membranes lining part of a duct, and each membrane is backed by a cavity filled with helium. For a typical configuration of a duct with height h, membrane length $L=7h$, cavity depth $h_c=0.2h$, and tension $T=0.52\rho_0 c_0^2 h^2$, where ρ_0 and c_0 are the ambient density and speed of sound in air, respectively, the transmission loss has a continuous stop band of TL$>$6.35 dB for frequency $0.03c_0/h$ to $0.064c_0/h$, which is much better than traditional duct lining. In addition to the mechanisms at work for drum silencers with air cavity, the low density of helium reduces the masslike reactance of the cavity on the second *in vacuo* mode of membrane vibration. The reduction greatly enhances the membrane response at this mode, which is found to be critical for achieving a broadband performance in the low-frequency regime. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1603232]

PACS numbers: 43.50.Gf, 43.20.Tb, 43.20.Ks [DKW]

I. INTRODUCTION

Low-frequency noise is difficult to control and it still presents a technical challenge. Engineering applications include ventilation systems and flow-through devices such as wind tunnels, engine testing facility, etc. There are two categories of noise control methods: one is passive and another active. Active noise control holds the promise for controlling low-frequency noise, but issues of reliability and cost remain. Passive control is still seen as the ultimate engineering solution. At present, passive control measures mainly rely on duct lining, which is not effective at low frequencies. It also has environmental problems as fibers exposed to flow trap dusts and it could become a health hazard.

Resonators and expansion chambers are the backbone of traditional design of broadband silencer. Different configurations of expansion chambers incorporating resonators have been studied analytically and experimentally. Sullivan (1979a, 1979b) studied, both theoretically and experimentally, the perforated tube muffler formed by one perforated tube enclosed by a chamber. In order to enhance the level of noise reduction and eliminate the passband, further studies on the combination of multiple perforated mufflers were carried out (Thawani and Jayaraman, 1983). There are three main types of multiple perforated tubing mufflers: concentric, plug muffler, and three-duct cross-flow muffler, and their performance are well verified by experiments (Munjal *et al.*, 1993). Mufflers with good performance, such as the plug muffler, tend to carry a penalty of high back pressure. To reduce the pressure drop, a design of two-chamber, three-duct, open-ended muffler is recommended by Gogate and Munjal (1995). It works for a wider frequency range and the pressure drop is not very large. However, it is larger than the concentric and plug type mufflers. As the pressure loss in the muffler connected to an engine or pump wastes power, more power is required for the power source, and this eventually intensifies the noise source. So, apart from the consideration of the environmental aspects and the space occupied, back-pressure is one of the most important attributes of a muffler.

The team led by Fuchs (2001a) has been succeeding in introducing many fiber-free solutions to engineering applications. The key technique is the use of microperforated sheets and other forms of Helmholtz resonators. A device called a membrane absorber box is one such example (Ackermann *et al.*, 1988). Its practical performance as an exhaust stack silencer has been reported by Ackermann and Fuchs (1989). This performance is discussed in some more detail in Sec. IV where comparison is made with the present method.

Recently, Huang (1999) introduced the idea of using tensioned flexible membranes to reflect low-frequency noise. This is a reactive noise control method. The objective is identical to that behind the development of many patented devices by Fuchs and his colleagues (1988, 2001a, 2001b), which is to provide a fiber-free solution. In addition to this, emphasis has been placed on the following two aspects: (a) total noninvasiveness with a view to applications where pressure loss is extremely important or any intrusion into the flow conduit by a muffler is very undesirable, and (b) minimal space required outside the main flow conduit. These objectives are achieved through the use of highly tensioned membranes backed by cavities, and theoretical predictions (Huang, 2002) are validated by experiments (Choy and Huang, 2002). Since the device resembles a drum, it is called a drum-like silencer, or drum silencer henceforth.

The exact mechanisms of a drum silencer have been revealed in detail in Huang (2002) for an idea configuration,

[a)]Author to whom correspondence should be addressed; electronic mail: mmlhuang@polyu.edu.hk

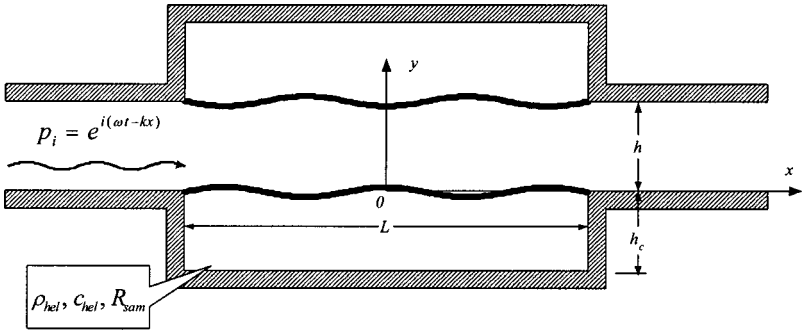


FIG. 1. Configuration model of sound incident on a stretched membrane backed by a cavity filled with helium gas.

and further studies were conducted by Choy (2003) which also takes into account realistic factors such as damping. The findings are now summarized before embarking on the study of helium filled cavities. When the membrane response is decomposed by *in vacuo* modes, it is found that the first two modes play the dominant role. The first mode is most effective in reflecting sound, but it is very difficult to excite due to the relative incompressibility of air inside the cavity. The second mode does not involve any change of cavity volume and is therefore easier to excite, but the radiation efficiency is low due to its dipole-like feature. The third and higher order modes are not very receptive to the incoming low-frequency sound, and they are ineffective in reflecting sound. It is shown that a high tension applied on the membrane promotes the response of the membrane in lower order modes, and the system becomes effective in reflecting sound. The spectra of sound reflection and transmission loss show many peaks at frequencies where sound is almost completely reflected. The performance of the membrane at frequencies between two adjacent peaks can be maintained at a rather high level when appropriate membrane properties are chosen. Damping is not particularly helpful in the enhancement of transmission loss in this case.

The performance of such device is assessed by the level of transmission loss as well as the frequency bandwidth. A rather aggressive criterion of transmission loss is used for the stop band. The lowest frequency which can be controlled in such a drum silencer is typically related to the quarter wavelength of the incident noise, hence related to the length of the membrane. The fundamental reason why long membranes are needed is because the radiation efficiency of the dipole-like second *in vacuo* mode is low for short membranes. This factor is imposed by the sound speed in the main flow duct, which cannot be modified. In order to limit the volume of the cavity, one naturally seeks the reduction of cavity depth. This study explores the use of low density gas like helium in a very shallow cavity, which can be implemented by enclosing the gas in a thin plastic bag, see Sec. III. The result shows that the low acoustic impedance of the gas compensates for the cavity shallowness and the transmission loss achieved is very encouraging. Section II introduces the theoretical prediction method and the optimization of membrane tension. Section III analyzes the results of an optimal configuration found in Sec. II, and Sec. IV compares the theoretical prediction with experimental results.

II. THEORY

Since the details of the basic solution method have been described in Huang (2002) and Choy and Huang (2002), only essential features of the solution and modifications needed for the current study are presented.

A. The solution method

The theoretical model is shown in Fig. 1. The geometry resembles a standard, two-dimensional expansion chamber with a main channel height h, and two identical cavities of length L and depth h_c. Two membranes of length L cover the cavities. With such a cover, the gas medium inside the cavities can differ from air, such as helium gas, and the gas density and speed of sound are denoted as ρ_{hel} and c_{hel} for helium. For convenience, the drum silencer in which the cavity is filled with normal air is denoted as drum-silencer-air, or DSA for short, while that filled with helium as DSH.

Before studying the membrane dynamics, a normalization scheme is introduced to identify controlling parameters. Three basic quantities are used for normalization: the main duct height h as the length scale, air density ρ_0, and the air speed of sound in free space, c_0. The normalized frequencies and the wave number are

$$\bar{f} = fh/c_0, \quad \bar{\omega} = \omega h/c_0 = 2\pi\bar{f}, \quad \bar{k} = kh. \tag{1}$$

The dimensionless first cut-on frequency of the duct is $\bar{f} = 0.5$. The membrane mass per unit area, M, is normalized by $\rho_0 h$, hence the mass ratio $\bar{M} = M/(\rho_0 h)$, and the tensile force T by $\rho_0 c_0^2 h^2$, and pressure by $\rho_0 c_0^2$. For the sake of brevity, all over-bars for dimensionless quantities are dropped henceforth.

When a harmonic sound of unit amplitude $p_i = e^{i(\omega t - k_0 x)}$, where $k_0 = \omega/c_0$ is incident on the membrane, the membrane responds by vibrating at a velocity V, which gives rise to a pressure difference across the two membrane surfaces. If the total loading on the membrane is written as $p_i + \Delta p$, the dynamics of the fluid-structure coupling is described by the following equation for the displacement calculated as $\eta = V/(i\omega)$:

$$M\frac{\partial^2 \eta}{\partial t^2} - T\frac{\partial^2 \eta}{\partial x^2} + \Delta p + p_i = 0. \tag{2}$$

The fluid loading term, Δp, can be divided into three parts. One is the pressure induced on the duct side of the membrane, denoted as $p_{+\mathrm{rad}}$, and the second one is on the cavity

side of the membrane, $p_{-\mathrm{rad}}$, when the cavity is first regarded as an infinite channel of height h_c. The third part, $p_{-\mathrm{ref}}$, is the reflected waves from the two vertical walls in the channel which are actually the end walls of the cavity. $p_{+\mathrm{rad}}$ can be found by the summation over all duct acoustics modes, ψ_m, as shown in the following (Doak, 1973) in dimensional form:

$$p_{+\mathrm{rad}}(x,y,t) = \frac{\rho_0}{2h}\sum_{m=0}^{\infty} c_m\psi_m(y/h)\int_{-L/2}^{+L/2}\psi_m(y'/h)V(x',y',t) \times[H(x-x')e^{-ik_m(x-x')}+H(x'-x)e^{+ik_m(x-x')}]dx', \tag{3}$$

$$\psi_m(y/h)=\sqrt{2-\delta_{0m}}\cos(m\pi y/h), \quad k_m=\frac{\omega}{c_m},$$

$$c_m=\frac{ic_0}{\sqrt{(m\pi/k_0h)^2-1}},$$

where c_m and k_m are the phase speed and wave number of the mth duct mode, respectively, H is the Heaviside function, and δ_{0m} is the Kronecker delta. The integration is carried out over the source surface on $y'=0$, $x'\in[-L/2,L/2]$. To find $p_{-\mathrm{rad}}$ in the lower cavity, ρ_0 and c_0 are replaced by the corresponding properties in the cavity, such as ρ_{hel} and c_{hel} for helium, the vibration velocity V replaced by $-V$, and h replaced by h_c. The reflection loading, $p_{-\mathrm{ref}}$, is found by expressing the standing wave pattern as a sum of duct acoustics expansion like Eq. (3) with two Heaviside functions replaced by two constant, A_m and B_m,

$$p_{-\mathrm{ref}}(x,y,t)=\frac{\rho_{\mathrm{hel}}}{2h_c}\sum_{m=0}^{\infty} c_{mc}\psi_m(y/h_c)\int_{-L/2}^{+L/2}\psi_m(y'/h_c) \times[-V(x',y',t)][A_me^{-ik_{mc}(x-x')} + B_me^{+ik_{mc}(x-x')}]dx', \tag{4}$$

$$c_{mc}=\frac{ic_{\mathrm{hel}}}{\sqrt{(m\pi/k_{\mathrm{hel}}h_c)^2-1}}, \quad k_{\mathrm{hel}}=\frac{\omega}{c_{\mathrm{hel}}}, \quad k_{mc}=\frac{\omega}{c_{mc}},$$

which are found by the rigid wall conditions at $x=\pm L/2$,

$$\left.\frac{\partial(p_{-\mathrm{rad}}+p_{-\mathrm{ref}})}{\partial x}\right|_{x=\pm L/2}=0,$$

$$\rightarrow A_m=\frac{e^{ik_{mc}(L-2x')}+1}{e^{ik_{mc}(2L)}-1},$$

$$B_m=\frac{e^{ik_{mc}(L+2x')}+1}{e^{ik_{mc}(2L)}-1}. \tag{5}$$

The Galerkin procedure is followed to solve Eq. (2). The velocity of the membrane, V, is expanded into a series of sine functions (*in vacuo* modes) for a membrane simply supported at the two edges, $|x|=L/2$,

$$V=\sum_{j=1}^{\infty} V_j\sin j\pi(x/L+1/2),$$

$$V_j=\frac{2}{L}\int_{-L/2}^{+L/2}V(x,t)\sin[j\pi(x/L+1/2)]dx, \tag{6}$$

and the fluid loading is expressed as a sum of modal contributions,

$$\Delta p=\sum_{n=1}^{\infty}\sum_{j=1}^{\infty} V_jZ_{jn}\sin[n\pi(x/L+1/2)], \tag{7}$$

where the single mode impedance Z_{jn} is defined as the nth fluid loading coefficient caused by the vibration of the jth *in vacuo* mode of unit amplitude. The response of the membrane vibration is found by solving the following set of coupled dynamic equations,

$$\mathcal{L}_iV_j+\sum_{n=1}^{N} V_nZ_{jn}=-I_j, \quad j=1,2,..., \tag{8}$$

where

$$I_j=\int_{-L/2}^{+L/2}\frac{2p_i}{L}\sin[j\pi(x/L+1/2)]dx = 2j\pi e^{ik_0L/2}\left[\frac{1-e^{i(-k_0L+j\pi)}}{(j\pi)^2-(k_0L)^2}\right] \tag{9}$$

is the modal coefficient of the incident wave (forcing), and

$$\mathcal{L}_j=Mi\omega-i\frac{T}{\omega}\left(\frac{j\pi}{L}\right)^2 \tag{10}$$

is the *in vacuo* linear operator or membrane reactance. N can be truncated to about 25 and standard matrix inversion techniques like Gaussian elimination can be used. Note that the problem of two membranes shown in Fig. 1 can be solved by splitting the main duct into two parts, each with one membrane and the duct center is replaced by a rigid wall.

Once the vibration velocity is found via the modal coefficient V_j, the reflected wave, denoted as p_r, can be found by evaluating the radiated wave into the far left from Eq. (3), $p_r=p_{+\mathrm{rad}}|_{x\rightarrow-\infty}$, while the transmitted wave, p_t, is found by the superposition of the incident wave, p_i, with the radiation into the far-right, $p_t=p_{+\mathrm{rad}}|_{x\rightarrow+\infty}+p_i$, where $|p_i|=1$. The coefficients of energy flux reflection β, and absorption α, and the transmission loss TL can be evaluated as follows:

$$\beta=|p_r|^2, \quad \alpha=1-\beta-|p_t|^2, \quad \mathrm{TL}=-20\log_{10}|p_t|. \tag{11}$$

The complex amplitude of the reflected sound, p_r, is the sum of contribution made by all individual membrane vibration modes, which is found as follows with the help of Eq. (3):

$$p_r=\frac{1}{2}\int_{-L/2}^{+L/2}V(x')e^{-ik_0x'}dx'=\sum_{j=0}^{\infty}V_jR_j,$$

$$R_j=\frac{1}{2}\int_{-L/2}^{+L/2}\sin[j\pi(x'/L+1/2)]e^{-ik_0x'}dx', \tag{12}$$

where R_j is the complex amplitude of the reflected sound by the induced vibration of the jth mode with unit amplitude.

For the convenience of later analysis, it is pointed out that there is no cross-modal interaction between even and odd modes, and the effects of higher order modes are small at low frequencies (Huang, 2002). As a result, the low-frequency performance can be analyzed in terms of the decomposed first and second mode solutions as

$$V_1=\frac{I_1}{\mathcal{L}_1+Z_{11}},\quad V_2=\frac{I_2}{\mathcal{L}_2+Z_{22}}. \tag{13}$$

B. Prediction of optimal parameters

In a parametric study conducted by Huang (2003), the criterion value of transmission loss to be qualified as stop band is given as follows:

$$\mathrm{TL}_{\mathrm{cr}}=10\log_{10}[1+\tfrac{1}{4}(a_r-a_r^{-1})^2],\quad a_r=1+\sqrt{3\Lambda}, \tag{14}$$

where $\Lambda=2h_cL$ is the total volume of the two cavities. This value is the peak transmission loss of an expansion chamber made of two rectangular cavities which, when put together, form a cubic (square in two dimension) cavity of volume three times the total volume occupied by the drum silencer cavities. The frequency range in which TL satisfies the condition of $\mathrm{TL}\geqslant\mathrm{TL}_{\mathrm{cr}}$ continuously is defined as the stop band, denoted as $[f_1,f_2]$. An optimization procedure is constructed to search for the best tension of the membrane that gives the highest frequency ratio f_2/f_1. In defining the objective function this way, due emphasis is placed on the low frequency noise control. The following set of parameters are used as the default values:

$$M=1.4,\quad h_c=0.2,\quad L=7. \tag{15}$$

The result for the membrane with different tensions and the optimum result are shown in Fig. 2. The performance of the helium filled cavity is compared with that of normal air cavity of the same size. For easier comparison with experimental data, dimensional tension is indicated in Fig. 2 for the test rig of square cross section of $h=100$ mm. Many stop bands are found, but our attention is on the one with the lowest frequency. For the default total volume of the cavity $\Lambda=2h_cL=2.8$, $\mathrm{TL}_{\mathrm{cr}}=6.35$ dB.

When the tension is zero, as shown in Fig. 2(a), there is only inertia effect in the structural property. The spectral pattern is similar to that of a simple expansion chamber when the membrane is removed from the helium cavity. By comparing the frequency ratio f_2/f_1, the dimensional optimum tension to achieve the best performance in the default configuration is found to be 750N with $f_2/f_1=2.17$ as shown in Fig. 2(c). When the tension is lowered to 650N, the TL trough between the first and second peaks in Fig. 2(b) falls below $\mathrm{TL}_{\mathrm{cr}}$. When the tension is increased beyond 750N, the frequency ratio becomes smaller as the first and second peaks close up. f_2/f_1 is about 2.10 for 900N.

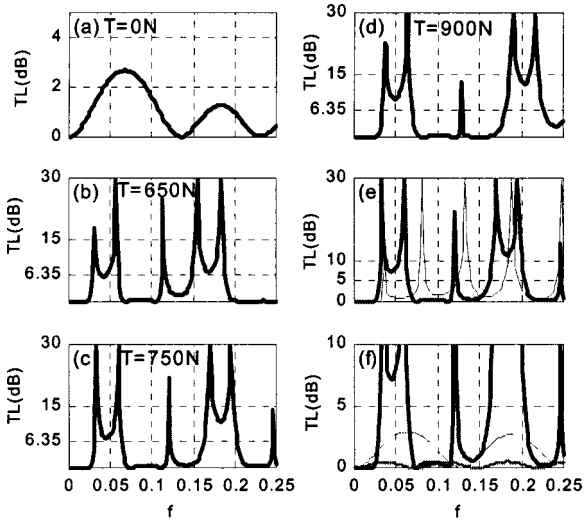


FIG. 2. The prediction of optimal DSH performance and its comparison with other reference configurations. (a), (b), (c), and (d) compare the TL for DSH with membrane tensions of $T=0$N, 650N, 750N(optimal) and 900N; (e) compares the optimal DSH (thick line) with DSA (thin line); (f) compares the optimal DSH (solid line) with expansion chamber with air in the cavity (*), and expansion chamber with helium in the cavity (---).

Having found the optimum tension, 750N, for the membranes backed by helium cavities, more detailed analysis can be carried out. First, comparison is made with the membranes in the air cavity with the same depth (DSA). Figure 2(e) compares the optimal DSH with 750N (thick line), and with DSA with 2870N (thin line). The latter represents the best possible TL pattern for DSA although the best is also mostly below $\mathrm{TL}_{\mathrm{cr}}=6.35$ dB. By now it is apparent that helium cavity can enhance the transmission loss at low frequency. Such performance is attributed to the fluid loading in the cavity and this will be further discussed in the following section. As shown in Fig. 2(f), even when the membrane is removed, the TL of the expansion chamber with helium cavity (---) is much better than that with air cavity (*). The first TL lobe of the helium expansion chamber covers two TL lobes of the air expansion chamber. The width of the lobe is characteristic of a normal expansion chamber with a length of $3.9h$ and the helium expansion chamber also gives much higher TL. This means that the helium cavity is equivalent to a deeper air cavity. Figure 2(f) also compares the DSH (thick line) with the helium expansion chamber, the addition of the membrane is found to enhance the TL level at low frequencies. However, at the very low frequencies ($f<0.027$), DSH degrades the performance of a helium expansion chamber. This will be further explained in the next section.

C. Analysis of the optimal helium filled drum silencer

Focusing on the low-frequency range of $f<0.15$, the spectra of the modal responses for the optimal DSH are shown in Fig. 3. Figure 3(a) shows the TL spectrum in which two peaks and two troughs are marked for membrane response analysis in the right column of Fig. 3. They are $P_1(f=0.0329)$, $P_2(f=0.0587)$, $T_1(f=0.0457)$, and $T_2(f=0.0906)$. Figures 3(b) and (c) analyze the vibration ampli-

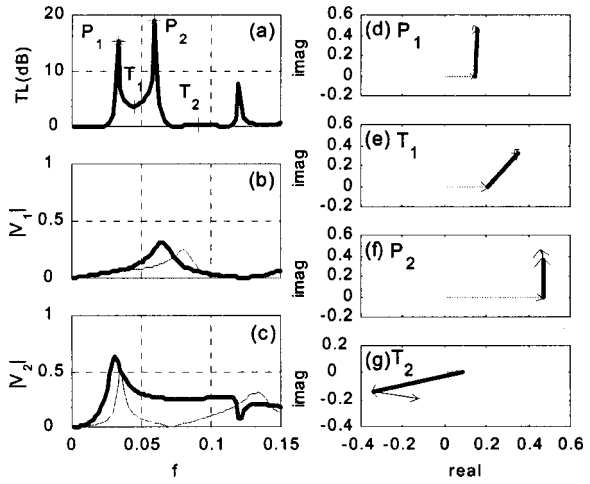


FIG. 3. Modal analysis of the optimal DSH. (a) The TL spectrum of DSH. (b) and (c) Comparison of the modal amplitudes of the first and second order modes of the optimal DSH (thick lines) and those of DSA (thin lines). (d), (e), (f), and (g) The complex modal contributions at the points of P_1, T_1, P_2, and T_2 marked in (a) are plotted.

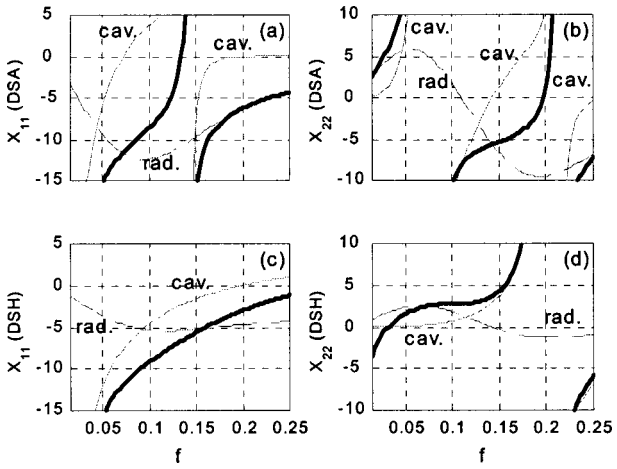


FIG. 4. Reactance of the first two modes of the optimal DSH at $T=750$N and DSA at $T=2870$N. (a) The components of the first mode reactance of DSA. The components are from radiation (marked ''rad''), cavity reflection (marked ''cav''), and the total reactance (thick line) including the structural property. (b) The components of the second mode reactance of DSA. (c) and (d) The first and second mode reactance of the optimal DSH are shown, respectively.

tudes of the first two modes, $|V_{1,2}|$, and compare the results (thick lines) with those of DSA at $T=2870$N (thin lines). In Fig. 3(b), it is observed that the peak of the first modal response of DSH is shifted to lower frequency but the amplitude remains almost identical to that of DSA. In contrast, the second modal response of DSH shown in Fig. 3(c) is greatly increased compared with DSA over a wide frequency range. It is also observed that the second modal response in DSH is higher than the first for the whole frequency range shown in Figs. 3(b) and (c). At about $f=0.065$, the first and second modal responses are almost the same and the levels are not small. This helps to enhance the transmission loss at this particular frequency. To understand the formation of the TL peaks and other frequencies marked in Fig. 3(a), the complex modal reflection contribution (V_jR_j) are shown in the subfigures in the right column as vectors. The first modal reflection (V_1R_1) is taken as the reference direction of real axis, and the second mode is highlighted by thick lines. Figure 3(d) depicts the composition of the modal reflections at the first peak (P_1) of $f=0.0329$. The magnitude of the first modal reflection (V_1R_1) is smaller than the second (V_2R_2). The sum of all the other modes has very small magnitude and it has little effect on the first peak (P_1). Therefore, it can be said that the first TL peak is mainly contributed by the second mode. Figure 3(f) shows the modal reflection contributions for the second peak (P_2) at $f=0.0587$. The first vector is longer than the second, and the two are almost perpendicular to each other, hence no interference. The sum of the rest is again very small, meaning that the higher order modal reflections have almost no effect on the peak. It may be concluded that the second peak P_2 is mainly contributed by the first mode and the second mode is orthogonal to the first.

The point at frequency 0.0436, T_1, lies in between the first and second TL peaks, and is analyzed in Fig. 3(e). In this case, the first and the second modal reflection contributions are almost equally important and there is constructive interference. This is perhaps the reason why the TL dip at this frequency is not too serious. The magnitude of higher order modes is small. The second trough point at $f=0.0906$ is depicted in Fig. 3(g). V_1R_1 is extremely small, while V_2R_2 is relatively large, but it is canceled to a large extent by the sum of other modes. As a result, the transmission loss at this frequency is very small.

The frequency band between the first and second peaks has a desirable transmission loss. This is mainly caused by the dominance of the first and second modes from the above-presented vectorial analysis. The second mode makes larger contributions toward lower frequencies. The spectral variations of responses can be predicted mathematically from Eq. (13), and the analysis should focus on the modal impedance Z_{jn} and structural properties $\mathcal{L}_j$. Figure 4 shows the reactance of the first two modes of the membrane in the air cavity [top row, Figs. 4(a) and (b)], and helium cavity [bottom row, Figs. 4(c) and (d)]. The structural properties for the two cases are different because the optimal tensions are different. The curve marked ''cav'' indicates the cavity reflection effect from $p_{-\mathrm{ref}}$, while the curve marked ''rad'' is for the sum of the radiation from the membrane on the upper and lower sides, $p_{+\mathrm{rad}}+p_{-\mathrm{rad}}$. The thick line represents the total reactance where the structural inertia and stiffness $\mathcal{L}_j$ are also taken into account.

Figures 4(a) and (c) concern the reactance of the first mode for the two configurations. It is observed that the first mode is rather stiff for very low frequencies ($f<0.02$) mainly due to the cavity reflection effect, leading to very weak membrane responses for both DSH and DSA. For the optimal DSA, the cavity stiffness starts to decrease when the frequency goes beyond about 0.02, so the response shown in Fig. 3(a) starts to increase from frequency 0.02 until 0.079. At frequency about 0.144, there is almost no response of the first mode because the reactance diverges at the second cavity mode. This can be seen in Fig. 4(a) in which the total reactance (thick line) tends to infinity at this frequency. For the second modal reactance for DSA in Fig. 4(b), the cavity

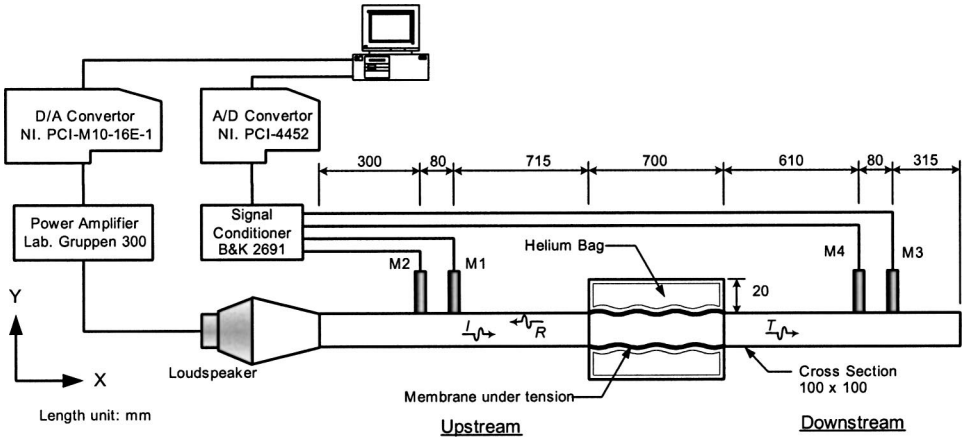


FIG. 5. The setup of the measurement system for DSH using the four-microphone, two-load method. Both data acquisition and the generation of the incident noise are controlled by the computer equipped with A/D and D/A cards, and a LABVIEW code.

modes' divergences occur at two frequencies within the range shown: $f=0.07$, 0.176. Such divergence seriously affects the response of the second mode, and V_2 vanishes at these frequencies. The highly oscillatory patterns of the total reactance shown in Figs. 4(a) and (b) are responsible for the many sharp peaks of TL for DSA in Fig. 2(e).

As shown in Figs. 4(c) and (d), for the optimal DSH, the divergence of the first and second modal reactance basically disappears except at frequency 0.2. It is also observed that the second modal reactance from cavity reflection, and radiation reactance are obviously decreased compared with DSA. The latter is very massive on the second mode for DSA with shallow cavity. The light gas helium effectively relieves the second modal cavity reactance. Therefore, the high response of the second mode can be kept over a wide range of frequency. The dip between the first and second peaks can also maintain a desirable level of TL. Figure 4(d) shows that the total reactance $X_{22}+\mathcal{L}_2$ diverges at frequency 0.2 where $V_2=0$. This represents a big shift from $f=0.07$ for the DSA shown in Fig. 4(b), and the shift provides good performance for low frequencies.

III. EXPERIMENTAL VALIDATION

Theoretical predictions show that the shallow DSH can control the low frequency duct noise by reflection. The stop band lies in the region of very low dimensionless frequencies. As frequency scales by c_0/h, noise of very low dimensional frequencies can be tackled if the cross section of the duct is large. Theoretical predictions like these need to be validated by experiment. Since any experimental rig is three dimensional in nature, the design to simulate the two-dimensional theoretical model should be conducted carefully. The crucial parts of the rig include the tension gear to apply tension uniformly across the third dimension, which is not included in the theoretical model, and the method to measure tension as well as the injection system of the helium gas without leakage from the test rig. The tests give transmission loss, reflection coefficient, and absorption coefficient.

Figure 5 shows the schema of the experimental setup (with dimensions labeled). The duct wall is made of 15-mm-thick acrylic, which is believed to be acoustically rigid. The first cut-on frequency of the duct is 1700 Hz. Two pairs of B&K $\frac{1}{2}$ in. condenser-type, intensity microphones (type number 4187) are installed flush with the duct walls. A wide separation distance of 80 mm is used for the microphone pairs in order to have a good measurement accuracy at low frequencies. The microphones are supported by a B&K's Nexus four-channel conditioning amplifier (type 2691), and the signals are acquired through the National Instruments' AD conversion card type PCI-4452. The noise is generated by a loudspeaker driven by DA signals. Both the AD and the DA processes are controlled by a LABVIEW program, which is made to run through a range of testing frequencies from 20 to 1000 Hz in a loop with a frequency interval of 10 Hz. The output signal from the DA converter (PCI-M10-16E-1) is passed to the loudspeaker via a B&K's power amplifier (LAB Gruppen 300). The discrete frequency approach is preferred. The natural advantage of the pure tone tests is the better signal-to-noise ratio, which is especially important at low frequencies where the loudspeaker is not quite effective. In addition, the sampling rate and sampling period can be adjusted automatically in the LABVIEW code such that an integer number of cycles is guaranteed for all frequencies tested. No digital windowing is needed for the subsequent fast Fourier transform analyses.

The exact locations of the microphones are shown in Fig. 5 with the length unit of mm. The wave components in the upstream side and downstream side indicated in Fig. 5 can be resolved by one pair of microphones on each side. The first test is conducted by putting a rigid plate at the termination of the downstream side. Another set of measurements was taken by replacing the rigid ending plate with a 500-mm-long tube filled with sound absorption materials providing a partially anechoic condition. The linear combination of the intensity fluxes for the two tests then gives a virtual test result in which the downstream is strictly anechoic. This is called the two-load method (Munjal and Doige, 1990). Before doing the experiment, the microphones and the system are calibrated and the procedure can be referred to (Choy, 2001).

In order to keep the helium in the cavity separate from the main duct, a thin plastic membrane bag is used to contain the gas in an air-tight manner. The bag is called, respectively, air bag or helium bag when air or helium is filled into the bag. A cylinder of helium gas with 99% purity produced by a Hong Kong gas company was used. The density and sound speed of helium gas or other gases are obtained from (In-

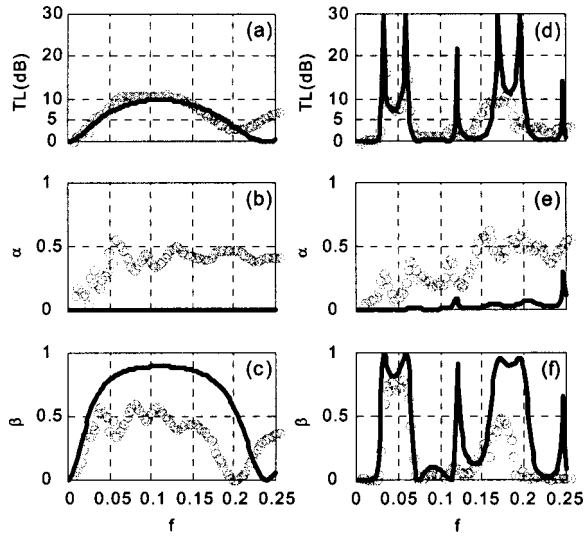


FIG. 6. Comparison of the experimental results (○) with theoretical predictions (—). The left-hand column shows the expansion chamber with helium cavity, but without membrane. The depth and length of the cavity are h and $5h$, respectively. The right-hand column is for the optimal DSH in the default configuration shown in Fig. 5.

cropera and Dewitt, 1996). The two cavities are 20 mm in depth. The plastic bag is made to a rectangular shape to fit into the cavity of 700 mm long and 20 mm deep. The mass per unit area of the plastic is 84 g/m^2.

The membranes are clamped and stretched by a tensile gear, for which the design details are given in Choy (2003) and briefly described here. The upstream end of the membrane is fixed by wrapping it around a cylinder of 12 mm in diameter, while the downstream end of the membrane runs through a 0.1 mm slit, and is fixed by two plates outside the duct. This end can be adjusted easily by the screw which fixes the tension. The tensile force applied is measured by a strain gauge glued to the surface of the membrane in the test section if there is no helium bag.

The stainless steel foil 0.025 mm thick is used in the experiment. The membrane is weighed, and the membrane mass per unit area is 0.17 kg/m^2. The mass ratio is $M/(\rho_0 h)=0.17/(1.225\times 0.1)=1.4$, a value which is used in Eq. (15).

The sub-figures in the left column of Fig. 6 give the experimental results of the helium expansion chamber. There is no metal membrane but the surface of the plastic bag is regarded as the plastic membrane. The dimension of this expansion chamber is not the same as the default values set in Eq. (15). The depth of the cavity of this helium chamber is 100 mm, so the area ratio of the expansion chamber to the duct is 3. The length of the chamber is 500 mm. The open circles represent the experimental result while the solid line is the theoretical prediction. Experimental results of TL in Fig. 6(a) show that the frequency range covered by the first lobe for the helium bag is smaller than prediction. The experimental results can be said to agree quite well with the theory below a frequency of 0.147. In between frequency 0.147 and 0.221, the theory overestimates. Since the maximum TL for a pure expansion chamber of an area ratio of 3 is only 4.4 dB, helium is seen to improve TL and the width of stop band. In Fig. 6(b), the energy loss in the experiment is found to be 45% on average, which takes away the contribution of the reflection coefficient shown in Fig. 6(c). Nevertheless, the main contribution to TL still derives from reflection, and the overall TL pattern in Fig. 6(a) still follows that of β shown in Fig. 6(c) instead of α shown in Fig. 6(b). The reason for the excessive level of α found in experiment is suspected to be the damping of the plastic bag which is under some tension during its inflation by helium. The maximum TL in the experimental result for the helium expansion chamber is found to be 10.6 dB and the peak is not very sharp. The trough goes down to 3 dB instead of 0 dB. The main finding is that both experiments and theory prove that helium can increase the effective depth of the cavity.

The right-hand column in Fig. 6 is for the metal membrane with helium cavity, or DSH, at the optimal condition of $T=750$N. Although there is metal membrane to separate two media in the cavity and the main duct, when the metal membrane is installed, the process of sucking the air out before the injection of helium tends to deform the metal membrane. Therefore, the plastic bag is still employed. Two remarkable peaks are found in the TL spectrum shown in Fig. 6(d). All predicted features are validated by the experimental data shown in open circles, although the peaks found in the experiment are smeared to lower values. The reason for the smearing is revealed by the plot of sound energy absorption coefficient α in Fig. 6(e). The experimental result basically matches the theoretical prediction based on an assumption of material loss factor for stainless steel $\sigma_1=0.5\%$ in the following equation, which is explained in the next section:

$$M(1-i\sigma_1)\frac{\partial^2\eta}{\partial t^2}-T\frac{\partial^2\eta}{\partial x^2}+\Delta p+p_i=0, \tag{16}$$

except at high frequencies ($f>0.147$) where the pattern of experimental data has shifted to lower frequencies. The corresponding stop band is frequency 0.03–0.064 and $f_2/f_1=2.13$ for the experimental results, while the theoretical prediction is 2.16. Notice that the transmission loss between the second and third peaks is very small, a passband, and this band covers a very wide frequency range. The damping of the plastic bag and the friction between the surface of the plastic bag and the rigid wall of the cavity may have caused the deviation between the experiment and theory at higher frequencies beyond the third peak. The absorption energy of the system is about 30% on average and this also causes the pattern of the reflection coefficient to deviate from the theoretical prediction. The main conclusion is that the performance of the DSH with shallow cavity can achieve good performance for the low frequency noise abatement. The energy damping found in helium experiment is further discussed in the following section.

IV. DISCUSSIONS

A. Modeling of the damping mechanism

The experimental data in Fig. 6(b) mean that the energy loss is mainly due to the effect of the plastic bag as there is no metal membrane. The vibration of the plastic bag causes friction with the rigid wall. This results in sound absorption

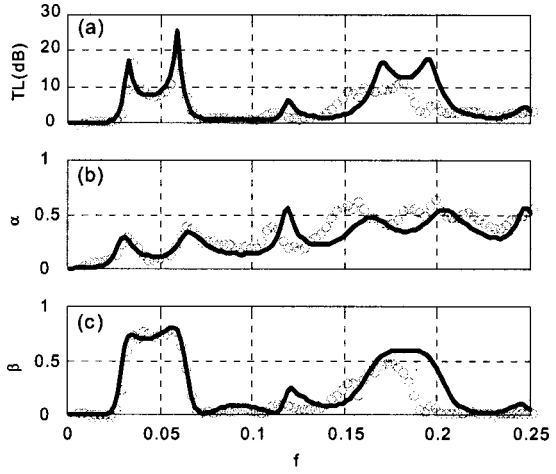


FIG. 7. Comparison of the experimental data (○) with the theoretical damping modeling (—) using $\sigma_1=2\%$, $\sigma_2=2.5\%$, and $R_{\text{sam}}=50\ \text{kg/m}^3\text{s}$.

in the cavity. This effect may be modeled by treating the cavity as one which is filled with some loose sound absorption materials of flow resistivity R_{sam}. This is called the sam-model. In addition, the metal membrane may also cause damping, and this is modeled by attaching damping terms in the following dynamic equation for harmonic oscillations of the fluid loaded membrane;

$$M(1-i\sigma_1)\frac{\partial^2\eta}{\partial t^2}-T(1+i\sigma_2)\frac{\partial^2\eta}{\partial x^2}+\Delta p+p_i=0. \qquad (17)$$

Details of all these damping models are given in Choy and Huang, (2002).

Theoretical tests show that a combination of three damping models with $\sigma_1=2\%$, $\sigma_2=2.5\%$, and $R_{\text{sam}}=50\ \text{kg/m}^3\text{s}$ give the best fit for the experimental data shown in Figs. 6(d), (e), and (f). The result is shown in Fig. 7. When the sam-model is used alone by setting $\sigma_1=\sigma_2=0$, it is difficult to match the first two peaks of α in Fig. 6(e) at $f=0.03$, 0.065. When R_{sam} is too large, it suppresses the first TL peak. The addition of the loss factor in the tensile force term, σ_2, produces better agreement between the theory and experiment for α. It also helps to match all the peaks in TL spectrum in Fig. 7(a). At the high end of the frequencies shown, the energy loss is very large, and this can only be modeled by the loss factor in the inertia term, σ_1. The combination of the three damping models manages to match the experimental results well although there is still some frequency shifting at high frequencies. This deviation may be caused by some uneven curvature of the plastic bag and the absorption in the third direction, which cannot be dealt with by two-dimensional theories.

B. Comparison with other devices

The drum silencer with the shallow helium cavity can help to enhance the noise reduction compared with the shallow air cavity. From the practical point of view, it should be taken to compare with some traditional silencer with the same configuration or dimension. Figure 8(a) shows the comparison of the DSH with duct lining using dimensional frequency as abscissa. The thickness and length of the duct

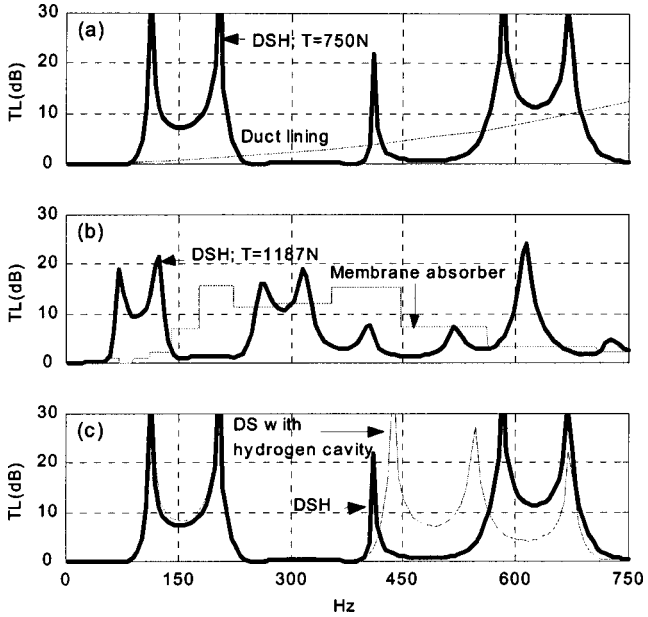


FIG. 8. Comparison between (a) DSH and duct lining, (b) membrane absorber and equivalent DSH, and (c) drum silencers with helium cavity and hydrogen cavity. (a) and (c) are for the default configuration, while (b) is done for the membrane absorber configuration.

lining used are 20 and 700 mm, respectively. The performance of the duct lining is evaluated by the method of Ingard [(1994); p. 6-6, Fig. 6.1.2. with 70% open area and $R=4$]. The DSH is better than the duct lining at low frequencies as such a shallow lining cannot absorb much low frequency noise. The duct lining always causes the hygiene problem to the environment, so the fiber-free DSH is an attractive alternative. However, the DSH performance at the middle frequency range of 200–400 Hz is not desirable.

Figure 8(b) shows the comparison of DSH with the membrane absorber employed as a splitter silencer for which a one-third octave band performance is given in (Ackermann and Fuchs, 1989). The membrane absorber is 1000 mm long and 100 mm wide. The air passage width is 150 mm. Prediction is made for a DSH which is also 1000 mm long and the cavity has a depth of 50 mm on each side of a duct unit of 150 mm in height. Note that a DSH of this dimension may not be the best shape for the given total cavity volume but, nevertheless, the prediction is made for the purpose of simple comparison. As some sound is reflected at the entrance to the splitter and such reflection is not included in the present prediction of DSH performance, the actual sound energy flux entering the splitter region is counted as the true incident wave when the result of Ackermann and Fuchs (1989) is modified for comparison in Fig. 8(b). The contribution of the reflection at the exit of the splitter is not excluded as it is assumed that all reflected sound is absorbed by the membrane absorber. It is observed that the DSH is better than the membrane absorber in the frequency range from 60 to 140 Hz, but, as expected, much worse in the middle passband range.

Figure 8(c) compares the DSH with a drum silencer with cavity filled by hydrogen, which has an even lower density than helium. Predictions are made for the two silencers with the same default geometrical configuration. If there is no

metal membrane covering the hydrogen enclosed by a light plastic bag, the TL peak can be better than that of the helium cavity because the characteristic impedance of hydrogen is even smaller than helium. When the tensioned membranes are added, the optimum tension is found to be 714N. The TL spectrum is generally similar to that of DSH at low frequencies. As helium is chemically more stable than hydrogen, the inert gas of helium is preferred.

C. Practical issues for the use of helium and membranes

Several issues related to the actual implementation of the concept of DSH are discussed here based on what is learnt from the two-dimensional modeling and experiments. The first issue is the significance of the plastic bag. In order to separate two media between the cavity and the main duct, the plastic bag is used, but it should be nonpermeable, and the material should be light enough to allow sound waves to pass through. The second issue is the selection of the material of the metal membrane, which might be exposed to different flow situations. The stress level for the DSH used in this study is around 300 MPa, which falls within the range of fatigue stress of stainless steel, which is around 383 MPa. In order to make sure that no fatigue occurs, good quality material should be chosen. If this DSH is employed in the vehicle exhaust system, the high temperature will cause a vast variation of tension of the metal membrane and this would definitely results in an undesirable performance. The third possible issue is flow-induced vibration. From the extensive studies by Huang (1998, 2001), it seems unlikely that the membrane would experience any form of flutter when the flow speed is well below the *in vacuo* tensile wave speed in the membrane. For the thin stainless membrane used,

$$c_T = \sqrt{\frac{T}{Mh}} = \sqrt{\frac{750}{6860\times(0.025\times10^{-3})\times0.1}} = 209.1 \ \text{m/s}.$$

This speed is unlikely to be approached in ordinary ventilation system. However, the problem of self-induced vibration should be more seriously considered if it is used in a high-speed wind tunnel.

V. CONCLUSIONS

The study of full coupling between the sound and membrane with the helium cavity has been carried out through mathematical analysis, physical explanation, and experimental study. The fluid loading by the helium cavity on the membrane and the modal response have been investigated. These help to understand the transmission loss pattern in detail. The following conclusions are drawn for the performance of the DSH with shallow cavities.

(1) Helium provides significant performance enhancement when used in very shallow cavities. With an appropriate tension on the membranes, a very wide stop band can be obtained. This is attributed to the dominance of the first two modes which have higher reflection capability at low frequencies than other higher order modes. The feature of very shallow cavity might be very useful for flow-through silencers with very limited sideway space.

(2) The concept of drum silencer allows the use of different gas in cavities, while a plastic membrane gives the practical realization method to contain the gas.

(3) The use of the light gas liberates the massive fluid loading on the second *in vacuo* mode of the membrane response, while there is little effect on the first order mode. A helium filled cavity is found to increase the response of the second mode substantially, so it helps to increase the noise reduction at low frequencies.

(4) For the typical configuration discussed here, the results represent a viable alternative to duct lining and other fiber-free solutions while maintaining the added advantage of zero backpressure.

(5) The theoretical predictions of the performance of DSH are validated by experimental data despite two major limitations of the theory. One is the two-dimensional nature of the theoretical model, and the other is the lack of knowledge of the damping mechanisms for which only *ad hoc* models are attempted.

(6) Practical issues of DSH have been discussed. It is believed that DSH can be used in the normal air conditioning system, but problems are anticipated in the condition of large temperature variations such as vehicle exhaust.

ACKNOWLEDGMENT

The authors thank the Hong Kong Polytechnic University for its support through various research projects (Nos. G-V618, G-YC 93).

Ackermann, U., and Fuchs, H. V. (**1989**). ''Technical note: Noise reduction in an exhaust stack of a papermill,'' Noise Control Eng. J. **33**, 52–60.

Ackermann, U., Fuchs, H. V., and Rambausek, N. (**1988**). ''Sound absorbers of a novel membrane construction,'' Appl. Acoust. **25**, 197–215.

Choy, Y. S. (**2001**). ''Experimental study of absorption and reflection of grazing sound by flexible panels,'' Proceedings of the Eighth International Congress on Sound Vibrations, Hong Kong, China, 2–6 July, pp. 1003–1010.

Choy, Y. S. (**2003**). ''Sound induced vibration and duct noise control,'' Ph.D. thesis, The Hong Kong Polytechnic University, Hong Kong, China.

Choy, Y. S., and Huang, L. (**2002**). ''Experimental studies of Drum-like silencer,'' J. Acoust. Soc. Am. **112**, 2026–2035.

Doak, P. E. (**1973**). ''Excitation, transmission and radiation of sound from source distribution in hard-walled ducts of finite length. I. The effects of duct cross-section geometry and source distribution space-time pattern,'' J. Sound Vib. **31**, 1–72.

Fuchs, H. V. (**2001a**). ''Alternative fibreless absorbers—new tool and materials for noise control and acoustic comfort,'' Acta Acustica **87**, 414–422.

Fuchs, H. V. (**2001b**). ''Technical and applied papers—From advanced acoustic research to novel silencing procedures and innovative sound treatments,'' Acta Acustica **87**, 407–413.

Fuchs, H. V., Ackermann, U., and Rambausek, N. (**1988**). ''Sound attenuating box,'' United States Patent, US 4 787 473.

Gogate, G. R., and Munjal, M. L. (**1995**). ''Analytical and experimental aeroacoustic studies of open-ended three-duct perforated elements used in muffler,'' J. Acoust. Soc. Am. **97**, 2919–2927.

Huang, L. (**1999**). ''A theoretical study of duct noise control be flexible panels,'' J. Acoust. Soc. Am. **106**, 1801–1809.

Huang, L. (**1998**). ''Reversal of the Bernoulli effect and channel flutter,'' J. Fluids Struct. **12**, 131–151.

Huang, L. (**2001**). ''Viscous flutter of a finite elastic membrane in Poiseuille flow,'' J. Fluids Struct. **15**, 1061–1088.

Huang, L. (**2002**). ''Modal analysis of a Drum-like silencer,'' J. Acoust. Soc. Am. **112**, 2014–2025.

Huang, L. (**2003**). ''Parametric study of a drum-like silencer,'' J. Sound Vib. (unpublished).

Incropera, F. P., and Dewitt, D. P. (**1996**). *Fundamentals of Heat and Mass Transfer* (Wiley, Canada).
Ingard, K. U. (**1994**). *Note on Sound Absorption Technology* (Noise Control Foundation, U.S.A.).
Munjal, M. L., Krishnan, S., and Reddy, M. M. (**1993**). ''Flow-acoustic performance of perforated element mufflers with application to design,'' Noise Control Eng. J. **3**, 159–167.
Munjal, M. L., and Doige, A. G. (**1990**). ''Theory of a two source-location method for direct experimental evaluation of the four-pole parameters of an aeroacoustic system,'' J. Sound. Vib. **141**, 323–333.
Sullivan, J. W. (**1979a**). ''A method for modeling perforated tube muffler components,'' J. Acoust. Soc. Am. **66**, 772–778.
Sullivan, J. W. (**1979b**). ''A method for modeling perforated tube muffler components. II. Application,'' J. Acoust. Soc. Am. **66**, 779–788.
Thawani, P. T., and Jayaraman, K. (**1983**). ''Modeling and application of straight-through resonators,'' J. Acoust. Soc. Am. **73**, 1387–1389.

A fuzzy rule based framework for noise annoyance modeling

Dick Botteldooren[a] and Andy Verkeyn
Acoustics Group, Department of Information Technology, Ghent University, St. Pietersnieuwtraat 41, 9000 Gent, Belgium

Peter Lercher
Institute of Hygiene and Social Medicine, University of Innsbruck, Sonnenburgstrasse 16, A-6020 Innsbruck, Austria

(Received 13 December 2002; revised 20 June 2003; accepted 8 July 2003)

Predicting the effect of noise on individual people and small groups is an extremely difficult task due to the influence of a multitude of factors that vary from person to person and from context to context. Moreover, noise annoyance is inherently a vague concept. That is why, in this paper, it is argued that noise annoyance models should identify a fuzzy set of possible effects rather than seek a very accurate crisp prediction. Fuzzy rule based models seem ideal candidates for this task. This paper provides the theoretical background for building these models. Existing empirical knowledge is used to extract a few typical rules that allow making the model more specific for small groups of individuals. The resulting model is tested on two large-scale social surveys augmented with exposure simulations. The testing demonstrates how this new way of thinking about noise effect modeling can be used in practice both in management support as a ''noise annoyance adviser'' and in social science for testing hypotheses such as the effect of noise sensitivity or the degree of urbanization. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1604125]

PACS numbers: 43.50.Qp, 43.50.Rq, 43.66.Lj [DKW]

I. INTRODUCTION

Reported general annoyance is often used in social surveys to quantify the state of the acoustic environment in the vicinity of important sources.[1] Combined with noise level calculations, general dose-response relations[2–4] have been quite useful to characterize the annoyance experience of large populations. However, the value of such methods for predicting the effect of change[5–8] and for predicting individual response to noise of small groups is very limited.[9–11] In this paper the latter case is addressed. Annoyance is, in itself, a vague or fuzzy concept. It is only after averaging out over a large group of people that sharp statements may be made. Part of this vagueness is acknowledged in noise surveys by using words[12,13] to describe the degree of annoyance. Natural language is indeed by far a better tool to express perceptions than numbers. In everyday live, natural language is used to include fuzziness in evaluation quite often, e.g., the weather is *nice*; the tea is *cold*; this boy is *quite tall* for his age.

In this work natural language, or at least a very limited vocabulary, will be used to represent field knowledge on the construct of noise annoyance. But, even if granules or words are used to describe annoyance levels, this does not change the problem of predicting people's reaction to noise significantly. The prediction is still assumed to produce a single, crisp output, now being a word from the limited vocabulary. A solution can be found in allowing the model to predict the possibility that people use a particular word for describing their perception of the noise situation. It may well be that the model results in several words being equally possible answers. Adding additional input to the model and stuffing in more domain knowledge may make this model more decisive. However, a certain amount of inherent vagueness that can not be removed will always remain no matter how careful every little detail is modeled.

In the next section noise annoyance as a fuzzy concept will be discussed in more detail. Section III introduces fuzzy rule bases as a particular model to represent field knowledge. The next section then gives the general layout of the model, including the knowledge that is introduced. Finally, Sec. V is devoted to tuning the model and testing it on survey data.

II. NOISE ANNOYANCE AS A FUZZY CONCEPT

In the multidimensional universe of the state of mind or the general feeling of a person, there is a region that is called annoyance by noise. It is clear that this region cannot be limited by a crisp interval of measurable physical or physiological quantities. Nevertheless, it is possible to refer to it using natural language and we can clearly imagine how this state may feel or be like. It is even possible to communicate about it using questions like ''Are you annoyed by noise?'' and to answer by using words like ''Fairly.'' This is typical for a fuzzy concept. It is lying out there, somewhere; it can be referred to and talked about, but we cannot quantify it using crisp numbers, not even if intervals are allowed.

Consider now a universe U that has sufficient degrees of freedom to fully represent noise annoyance. This universe can be seen as an envelope of everything that is called noise annoyance. In the most general terms, the boundaries of U can be fuzzy. However, for simplicity they are considered to be crisp here. In this way, the concept is actually made crisp, but if the boundaries are set far enough, this has no influence on further treatment or development. Levels of noise annoy-

[a]Electronic mail: dick.botteldooren@intec.rug.ac.be

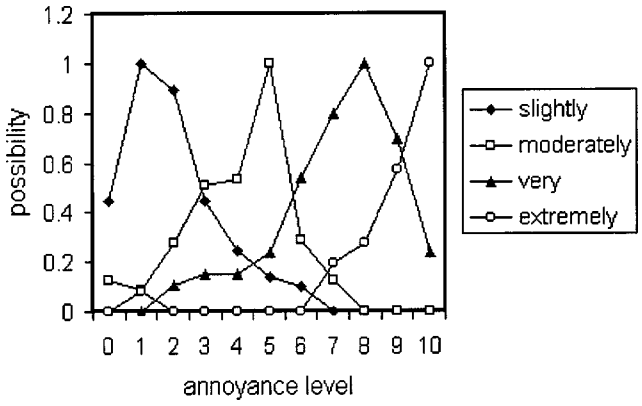


FIG. 1. Membership function for four noise annoyance labels.

ance can be identified as subsets L of U. It is clear that these L will be fuzzy sets because some elements of U are neither inside nor outside L.

In mathematical terms a fuzzy set is represented by a mapping $U \rightarrow [0,1]$ called the membership function, which can be seen as a direct generalization of the characteristic function $U \rightarrow \{0,1\}$ for classical sets. The fuzzy set can be referred to using a linguistic term, so, by extension, the membership function will be called the membership function of the linguistic term. A whole suite of operators (mainly generalizations of operators known from classical logic) is then available to perform operations on these membership functions.[14] Similarity measures can be constructed to compare fuzzy sets, and thus the linguistic terms used to describe them.[15]

The easiest way to picture the universe U is to limit it to one dimension and to represent it by a line segment of finite length. General techniques proposed in the fuzzy set literature can then be used to extract membership functions for linguistic terms used as labels of fuzzy sets subdividing U. In Ref. 12 these membership functions are extracted and compared for several languages based on the data gathered in an international annoyance scaling study.[13] Figure 1 shows the membership functions for four fuzzy sets and their associated labels.

III. FUZZY RULE BASE

In rule based fuzzy logic systems the domain knowledge is incorporated in a knowledge base that is formally described as a combination of a database and a rule base (Fig. 2). The database includes the definitions of the words and concepts used in the formulation of rules. These definitions can be extracted from observations using fuzzy clustering,

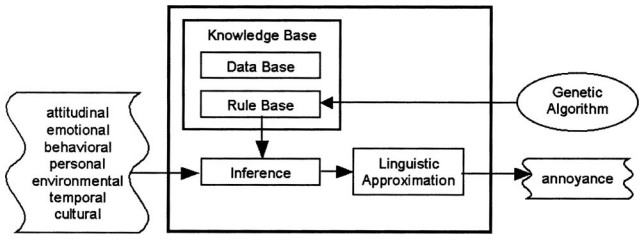


FIG. 2. The rule-based fuzzy logic system, showing its components.

they can be articulated using studies specifically aimed at finding the meaning of words, or experts can forward them. In this work, all words and concepts are defined on a one-dimensional universe. Their definition is therefore a one-dimensional membership function $\pi_L(u)$. By using type 2 fuzzy sets,[16] uncertainty on the meaning of words can be included. In this paper we refrain from doing so, mainly not to complicate things more than necessary at this point where fuzzy rule bases are used for the first time to model noise annoyance.

The rule base contains the logic modus ponens[17] type relations between fuzzy variables. A common inference system is used to implement the modus ponens. Assuming a rule "*IF X is A THEN Y is B*" and some vague knowledge "*X is A′*" are available, then the membership function B' for Y can be extracted. B' is calculated with the compositional rule of inference:

$$(\forall v \in V)(B'(v) = \sup_{u \in U} \min(A'(u), R(u,v))), \tag{1}$$

where R is the representation of the rule. Note that, in contrast to classical logic reasoning, conclusions can be drawn even if A' is different from A. Depending on the operator used to construct the rule representation R, different types of fuzzy rules are obtained.[18,19] In the noise annoyance model, the rules are formulated with a possibility interpretation in mind: the antecedent (e.g., *noise level* is *high*) is required to make the consequent (e.g., *noise annoyance* is *moderate or high*) possible. Therefore, in accordance with the literature cited, R is modeled using a t-norm, in particular the min-norm,

$$(\forall u \in U, \forall v \in V)(R(u,v) = \min(A(u), B(v))). \tag{2}$$

The results of inference [Eq. (1)] based on different rules articulated between the same universes U and V are aggregated with the maximum conorm as is required when using a t-norm to model the implication.

Each set of rules operating on the same input variable thus results in an estimate, H^n, for the outcome of the FRB model: the noise annoyance fuzzy set H. Since each H^n can be interpreted as a possible range for H, the intersection of all H^n is used as a best estimate for H. As a model for the intersection, the product t-norm is used. Compared to the minimum t-norm, a product conserves more of the information in each of the individual sets. Each H^n can also be regarded as an expert opinion based on a particular set of input data. The final result is a consensus between experts. Consensus can be implemented using the product of membership degrees. The fuzzy set for noise annoyance is thus obtained as $H(u) = \Pi_n H^n(u)$.

Each input variable, through its set of rules, creates the possibility for the noise annoyance fuzzy set to span a range H^n of U. Clearly, each H^n will not influence the outcome H of the model to the same extent. This is incorporated in the FRB by adding a fuzzy sufficiency variable, S, to the antecedent of the rules. This antecedent becomes

$$(X \text{ is } A) \text{ AND } (S \text{ is } S_H), \tag{3}$$

where S_H is a degree of sufficiency considered high enough to fully trigger the rule. This approach is a direct extension of the implementation of truth-qualifying statements as proposed in Ref. 20. To simplify the implementation, the information in S is reduced to a single crisp number, λ. The membership function $\pi_S(w)$ is equal to λ within the support of S_H and 0 outside it. Given the operators chosen to implement the implication, Eq. (3) can now be translated to a simple IF–THEN structure provided the consequent B is replaced by B^* defined as

$$B^*(v) = \min_v(B(v), \lambda). \tag{4}$$

The sufficiency, λ, for each rule to create a possible range for H can be imposed or, as is the case in this paper, it can be used to tune the model on a particular data set.

In many applications of FRB the fuzziness is only used as a methodology inside of the model and the result is defuzzified before making it available to the user. Defuzzification mostly results in a crisp number, but alternatively one can decide to describe the model outcome using word(s) by mapping the resulting fuzzy set to a collection of available terms. This procedure is called linguistic approximation. Similarity measures, calculating the similarity between the fuzzy set at hand (here the noise annoyance fuzzy set H) and the fuzzy set representation of each available linguistic term, L, are commonly used for this purpose.[12] The term, L, that matches best—has the highest similarity—with the outcome, H, of the fuzzy rule base is then chosen as the final linguistic result of the model. The translation of H to L can be interpreted as a partial defuzzification since the resulting label is still represented as a (type 1) fuzzy set for mathematical purposes. In Sec. V A we will argue that because of the inherent fuzziness in the concept of noise annoyance, it may be preferred not to defuzzify results to this extent.

IV. THE MODEL

In this section a general view on what a fuzzy rule based noise annoyance model (FRB) should look like is presented. This section follows the idea that fuzzy systems are useful to model reasoning with words or concepts, an idea proposed by the founder of fuzzy logic, Lofti Zadeh. The FRB is seen as a processing unit that mimics human perception of the acoustic stimulus and makes the decision for reporting noise annoyance. This view may lead to some grandiose statements, but we become more modest in the following sections. We also remind the reader that the examples given in this section are merely there to illustrate how a hypothesis can be included in the model not as proven facts on themselves. Figure 3 shows the interactions of the FRB system with its environment. The primary input triggering the whole process is clearly the acoustic field. The term acoustic field is used here to stress that the acoustical characterization is not limited to averaging acoustical indices such as L_{dn} or L_{den}. Nor is it limited to a single point in time or space. It was shown that additional soundscape factors indeed influence annoyance reactions to noise.[11,21,22] Nonacoustic factors can be regarded as preconditioners or modifiers. It can be either internal or external factors. Examples of such factors are

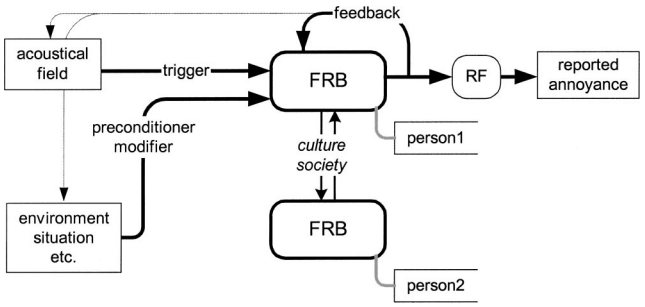


FIG. 3. Two interacting FRBs seen as a process surrounded by data flow (arrows), input and output interfaces (rectangles), and internal characteristics (open rectangles).

visual esthetics of the environment,[23–25] air pollution,[26–28] vibrations,[29–32] attitude towards the source,[10,33–35] and other ambient characteristics.[36–39] An external feedback mechanism that modifies reactions to noise on the basis of experienced noise annoyance has to be included. The positive influence of coping[40–43] is a possible example. For simplicity, the feedback is linked directly to the FRB representing the individual in Fig. 3. However, it can influence the acoustic field or the preconditioners and modifiers as well (dotted lines).

Interactions between individuals will lead to common interpretation and reaction to stimuli from the soundscape. The resulting tuning can be described as a form of culture or local subculture. In the FRB a primitive form of such a culture could possibly emerge by allowing the exchange of rules between FRBs. The current model does not include this, so a uniform ''culture'' is assumed over all subjects in the region to be modeled.

It is impossible to stick to a rigid similarity between the human individual and the FRB in further developing the model because the human is a complex system of traits and inner states that is influenced by many inputs from the past and even cross-individual inheritance. The modeling goal is therefore shifted from the individual to a cluster of individuals with similar behavior described by a set of indicators or factors. The indicators can be quite fundamental and closely linked to the traits or states they describe, but they can also be intermediates of other influences. As an example noise sensitivity is mentioned. This important trait variable cannot be described as triggered by a few external factors. However, indicators (e.g., number of children in household, age, number of rooms compared to number of people living in a house) could be found that determine clusters of people for which the noise sensitivity variable is higher than average in a fuzzy way.

Although three different groups of input variables are identified, acoustic field variables, preconditioners or modifiers, and clustering indicators, these variables are all treated very similarly from a mathematical point of view. In particular all of them can be crisp numbers or vague notions that may be described by words.

It turned out to be very useful to structure the FRB internally in a logical hierarchical way. Additional variables are introduced that are related to a number of input variables by fuzzy rules. If they can be measured using appropriate

survey questionnaires, then introducing these variables has the additional benefit that parts of the model can be tested separately.

The FRB is a framework that contains a set of rules that is constantly evolving as various aspects of noise annoyance are investigated. Moreover, different survey databases often include different sets of variables and although the FRB is perfectly handling missing data—since the fuzzy value ''undefined'' is always an option—removing or changing rules is clearly more efficient. Table I contains a typical set of rules included in the rule base together with the definitions of the linguistic terms that are used. For describing the membership functions, piecewise linear functions (PLFs) and more fluent S-shapes and bell-shapes are used. Figure 4 shows these membership functions for generic parameters specified in Table I.

It must be stressed that building this model is neither a data extraction exercise nor a brute force neural network like rule extraction from data. The rules mentioned above are based on expert opinion aggregated from literature. However, for some of the variables such as the population density in urban area, data clustering techniques are used to obtain the membership functions μ. Secondly, the rules presented here are to be interpreted as hypotheses. The tuning of the model based on survey databases, described in the next section, may well result in very low weights for some of the rules, making them almost obsolete. In contrast to other applications of FRB, the membership functions themselves, given in Table I, are not tuned.

V. TUNING OF THE MODEL

A. Quality measures

A classical crisp model is correct if it predicts the response of the individual exactly. The fuzzy noise annoyance concept starts from the idea that this is not possible. As described above, the fuzzy rule based system results in a fuzzy set H on the annoyance universe U. The word that best describes H can then be selected from a predefined language U_L (e.g., the five-point scale labels used in a survey). As explained in Sec. III, calculating the similarity between each of the word's membership functions and H, and simply selecting the word with the highest similarity to H, allows doing this.

However, the fuzzy set H represents not only the conceptual vagueness of annoyance but also the uncertainty or indecisiveness of the FRB system. As more information is fed to the system, H will become more precise. This means that each label, i, in the language, U_L, whose associated fuzzy set, L_i, is within H, is a possible description of the annoyance of the subject under test. If asked, the subject will ultimately select one of these labels. The FRB has no possible way of knowing which one. In the same way, any label, j, whose fuzzy set, L_j, is not in H, will not be a candidate descriptor for the annoyance experienced by the subject. Since containment of one fuzzy set in another is a fuzzy number, this results in two fuzzy sets on U_L: $P=\{X \in U_L | X \subseteq H\}$ and $N=\{X \in U_L | X \subseteq \neg H\}$, where $\neg$ is the NOT-operator. P contains all labels that are possible outcome of a survey; N contains those labels that are definitely not going to be used by any respondent. The membership degrees $\pi_P(i)$ and $\pi_N(i)$ of label i to the fuzzy sets P and N are defined as the degree of inclusion of the corresponding fuzzy set L_i in H and $\neg H$, respectively. They can be calculated as[44]

$$\begin{aligned} \pi_P(i) &= \mathrm{inc}(L_i, H) = \min_{\omega \in U}[\max(\mu_H(\omega), \mu_{\overline{L_i}}(\omega))], \\ \pi_N(i) &= \mathrm{inc}(L_i, \overline{H}) = \min_{\omega \in U}[\max(\mu_{\overline{H}}(\omega), \mu_{\overline{L_i}}(\omega))], \end{aligned} \tag{5}$$

where the line above a fuzzy set represents the complement of this set.

A first measure of quality of the model is a fuzzy extension of *false negative*. Assume that the model is tested by predicting the noise annoyance of a subject and that subject expresses its opinion selecting the label L_s. *False negative* then occurs when $L_s \in N$. In the fuzzy extension $\pi_N(s)$ is the degree to which *false negative* is true. Note that an equivalent definition of *false positive* cannot be used since H only states which annoyance levels are possible given the input to the FRB model. Complementary systems resulting in necessity quantification (where *false positive* would be a measure of quality) are an option, which is not envisaged here due to the significant increase in model complexity.

The previous measure of quality favors a system that is indecisive in a sense that not excluding any label from the result would be a good strategy. To make the system more useful indecisiveness must be included as an indication of poor quality. To measure indecisiveness a particular type of uncertainty, nonspecificity, as first introduced in Ref. 45 is used. It is defined as

$$NS(\pi_p) = \sum_{i=2}^{n} \pi_p(i) \log_2\left(\frac{i}{i-1}\right), \tag{6}$$

where the $\pi_P(i)$ have been put in decreasing order. Note that the above measure of quality only works with languages that are limited to a small number of labels that have membership functions with limited similarity between them (nonoverlapping membership functions).

B. Tuning methodology

The model is tuned by modifying the sufficiency of each rule in such a way as to minimize a suitable prediction error on a database obtained from a social survey combined with additional soundscape and demographic data. This optimization problem is solved using a genetic algorithm (GA).[46] These algorithms perform, in general, quite well on multivalley, noncontinuous fitness functions. Each individual in the population evolved by the GA is in this case an instance of the FRB model and is completely represented by a string of real values that is formed by the sufficiency of each rule in the FRB. The fitness of each individual—used by the GA in its imitation of natural selection—is the inverse of the prediction error of its FRB model on the database.

The choice of error measure determines the characteristics of the model that is obtained after tuning. If the fuzziness in the annoyance concept itself is neglected, then one may

TABLE I. Overview of a typical set of rules, expressing hypotheses, included in the FRB.

to model: annoyance			
IF L_{dn} (dBA)		THEN annoyance	
μ	label	μ	label
Trapezoidal(0,0,52.5,60)	*very low*	1-S shape(0,2)	
Triangular(52.5,57.5,65)	*low*	Bell shape(−3,1.5,5.5)	
Triangular(57.5,65,70)	*moderate*	Bell shape(1,4,9)	
Triangular(65,67.5,75)	*high*	Bell shape(5,8,11)	
Trapezoidal(67.5,75,80,80)	*very high*	S shape(7,10)	
IF distance to highway (m)		THEN annoyance	
μ	label	μ	label
Trapezoidal(0,0,1000,1500)	*close*	S shape(5,7.5)	*high*
Trapezoidal(1400,1600,inf,inf)	*far*	1-S shape(3.5,7.5)	*not high*
IF distance to main road (m)		THEN annoyance	
Trapezoidal(0,0,200,600)	*close*	S shape(5,7.5)	*high*
Trapezoidal(200,600,inf,inf)	*far*	1-S shape(3.5,7.5)	*not high*
IF distance to through road (m)		THEN annoyance	
Trapezoidal(0,0,60,250)	*close*	S shape(5,7.5)	*high*
Trapezoidal(60,250,inf,inf)	*far*	1-S shape(3.5,7.5)	*not high*
IF urban [center,city,suburb,country]		THEN annoyance	
μ	label	μ	label
Enum(0,0,0,1)	*country*	1-S shape(1.8,5.8)	*low*
Enum(0,0,1,0)	*suburb*	1-S shape(2.2,6.6)	*below moderate*
Enum(0,1,0,0)	*city*	S shape(2.2,7)	*above moderate*
Enum(1,0,0,0)	*center*	S shape(3,6.2)	*above moderate*
IF sensitivity to noise		THEN annoyance	
μ	label	μ	label
S shape(5,7.5)	*high*	S shape(0,2)	*at least somewhat*
1-S shape(0,4)	*not*	1-S shape(5,7.5)	*not high*
etc.			
to model: urban			
IF built up (% of area)		THEN urbanization [ce,ci,su,co]	
μ	label	μ	label
Trapezoidal(0,0,15,25)	*open*	Enum(0,0,0,1)	*country*
Triangular(15,25,30)	*around 25%*	Enum(0,0,1,0)	*suburb*
Triangular(25,30,35)	*around 30%*	Enum(0,1,0,0)	*city*
Trapezoidal(30,35,100,100)	*dense*	Enum(1,0,0,0)	*center*
IF population density (number/km^2)		THEN urbanization [ce,ci,su,co]	
μ	label	μ	label
Trapezoidal(0,0,770,1300)	*low*	Enum(0,0,0,1)	*country*
Triangular(1000,2200,3200)	*medium*	Enum(0,0,1,0)	*suburb*
Triangular(1300,2800,3800)	*medium*	Enum(0,1,0,0)	*city*
Trapezoidal(2200,3400,inf,inf)	*high*	Enum(1,0,0,0)	*center*
IF agriculture land (% of area)		THEN urbanization [ce,ci,su,co]	
μ	label	μ	label
Trapezoidal(0,0,15,25)	*little*	Enum(1,0,0,0)	*center*
Triangular(13,23,33)	*around 23%*	Enum(0,1,0,0)	*city*
Triangular(20,30,50)	*around 30%*	Enum(0,0,1,0)	*suburb*
Trapezoidal(30,50,100,100)	*a lot*	Enum(0,0,0,1)	*country*

TABLE I. (*Continued*.)

to model: sensitivity			
IF age (years)		THEN sensitivity to noise	
μ	label	μ	label
Trapezoidal(0,0,20,30)	*young*	1-S shape(5,7.5)	*not high*
Trapezoidal(50,60,100,100)	*old*	1-S shape(5,7.5)	*not high*
IF no. of children (no.)		THEN sensitivity to noise	
Triangular(0,0,2)	*none*	1-S shape(2,8)	*less than average*
Trapezoidal(0,1,2,4)	*few*	S shape(5,7.5)	*high*
Trapezoidal(2,5,10,10)	*many*	1-S shape(5,7.5)	*not high*
IF gender [male,female]		THEN sensitivity to noise	
Enum(0,1)	*female*	S shape(0,2)	*at least somewhat*

envisage a model that predicts a single label which equals the word chosen by a survey respondent, as often as possible. In that case, a similarity measure is used to obtain the label p whose membership function resembles H the most. The prediction error is obtained by counting incorrect prediction ($p \neq s$) in the database. Weighting with the occurrence of each annoyance level in the database improves predicting power in all ranges of U. Additional terms may be added to improve convergence or reduce nonspecificity.[47]

In an approach that fully acknowledges the inherent vagueness in noise annoyance, the measures of quality described in the previous section are combined to obtain the prediction error e:

$$e = \sum_{k=1}^{N} \frac{\pi_{N,k}(s_k)^2 + \alpha(NS(\pi_{P,k}))^2}{O(s_k)}. \tag{7}$$

The index k runs over all cases in the database and $O(s)$ is the relative occurrence of annoyance label s in the database. The weight α is introduced, attaching more importance to any of the two measures of quality. Depending on the value of weight α a more or less indecisive model is obtained, where the price to pay for decisiveness or specificity is more frequent false negative.

C. The databases

Two databases were used to tune and test the fuzzy noise annoyance model: a survey conducted in the Inn Valley in Tirol, Austria and a survey conducted in Flanders, Belgium.

A representative phone survey was conducted within an ongoing environmental health impact assessments of a new rail track in the Austrian part of the Alps near Innsbruck, which covers an area of about 40 km. This mainly rural area consists of small towns and villages with a mix of industrial,

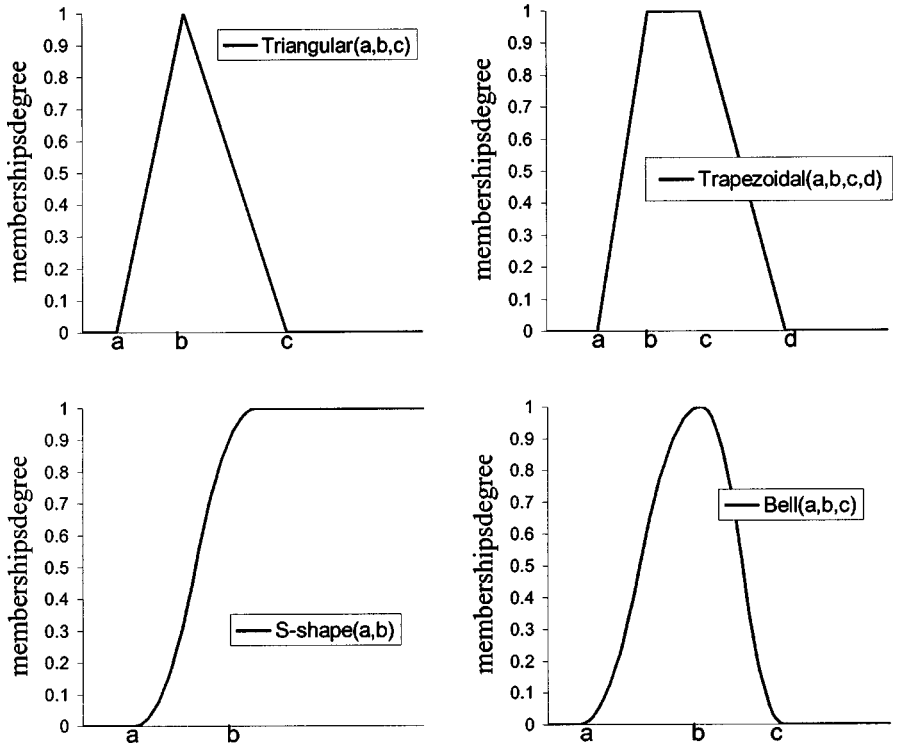


FIG. 4. Definition of membership functions used in the antecedent and consequent of the rules in the FRB model.

small business, and agricultural activities. The primary noise sources are road and rail traffic. In total, 2007 inhabitants were interviewed. The standardized interview (typical length 20 min) covered socio-demographic data, housing, satisfaction with public services and the environment, general annoyance, interference, coping with noise, and health. The overall response was 83%.

Subject selection was made from a Geographical Information System (GIS). Initially, 1500 inhabitants (aged 18 to 75) were sampled at random from the whole Inn-valley area (sample 1). This sample was enriched by another random sampling of 500 residents living within 150 m of the existing rail track and the highway or within 50 m of local roads (sample 2) to guarantee a sufficient number of people with higher exposure to noise and vibrations.

Noise exposure was assessed first by modeling (Soundplan) according to Austrian guidelines (ÖAL Nr 28+30, ÖNORM S 5011). Afterwards calibration was conducted and corrections were applied to the modeled data based on the recordings of 31 measuring stations. Based on both data sources approximate day–night levels (L_{dn}) were calculated for each respondent to ease comparison with typical dose-response data.

The social survey conducted in Flanders involved 3200 subjects. The general topic of the survey was the influence of odor, noise, and too much light on the living environment. The questionnaire covered socio-demographic data, housing, satisfaction with public services and the environment, general annoyance and annoyance caused by various sources, coping with noise and odor. The survey was presented as such to the subjects. The survey was part of the Investigation of the Environmental Living Quality performed on behalf of the Flemish Environmental Administration (AMINAL) by Deloitte&Touche and M.A.S. The overall response was 64%.

Road and rail traffic noise levels were calculated for each respondent's home. Exact rail-traffic and average noise emission of each type of train were used. For road traffic on major roads, simulated traffic intensity validated by results of several hundreds of counting stations were used. For the local roads an estimate of surface traffic for each geographic zone was made. The car and truck emission is taken from the recent revision of the Dutch guideline and propagation is calculated according to ISO 9613. Additional population and land use data were added.

D. Tuning results

Two important aspects of tuning must be highlighted. First, we will make clear how different flavors of FRBs can be obtained depending on the conditions imposed during tuning. Second, it is imperative to note that tuning has an influence on the effects of including more rules (hypotheses) in the model.

Depending on the importance attached to the FRB model being specific in its conclusions, another flavor of model can be obtained. Changing the weighting factor, α, in Eq. (7) allows implementing this during tuning.

This is illustrated using a road traffic noise annoyance model that includes rules based on L_{dn}, distance to the roads, degree of urbanization, reported traffic density, household size, age, and gender (Table I). Figure 5 shows the results of a model that is more precise, but often incorrect. Figure 6 shows the results of a model that is more fuzzy (has on the average more uncertainty in its output) but less often excludes the label chosen by the subject from its prediction. The area of the bubbles in these figures is proportional to the number of subjects in each category formed by a combination of nonspecificity and false negative. To illustrate more clearly the effect of tuning, the two clusters emerging in Fig. 5 are analyzed. In Table II the percentage of respondents in each of the five annoyance categories is given for the cluster with approximately zero *false negative* (not FN) and for the cluster with moderate nonspecificity and rather high *false negative* (not specific/FN). A combination of low nonspeci-

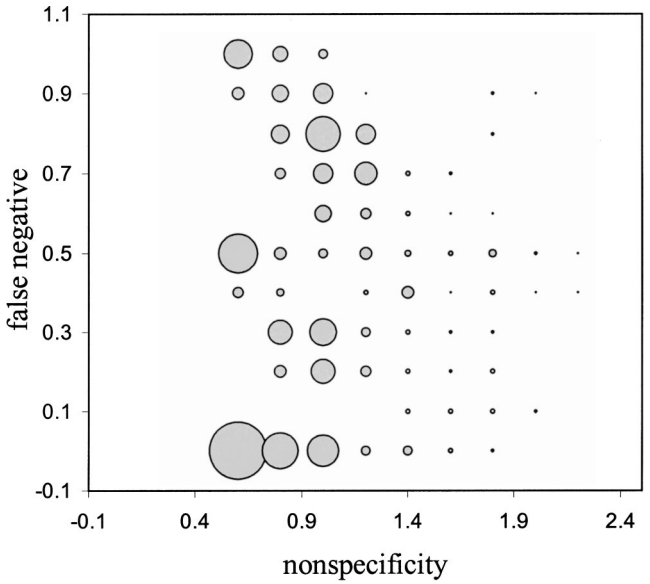


FIG. 5. Distribution of the subjects over a false negative versus nonspecificity plane for a model tuned with α=0.25; the area of the bubbles is proportional to the number of subjects.

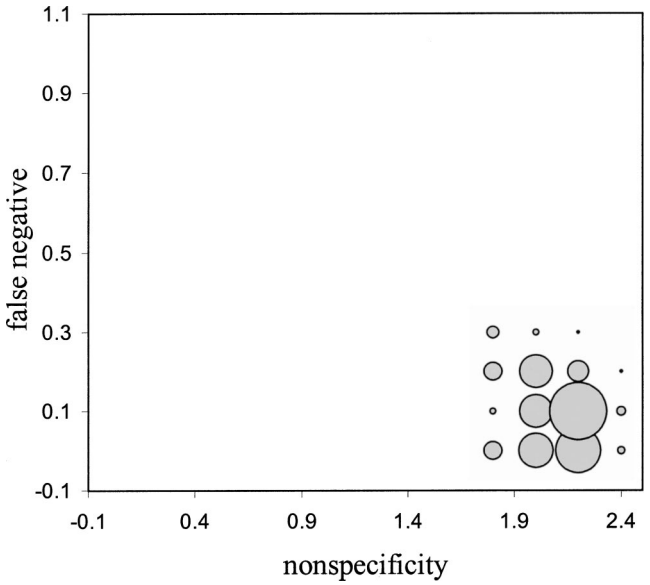


FIG. 6. Distribution of the subjects over a false negative versus nonspecificity plane for a model tuned with α=0.01; the area of the bubbles is proportional to the number of subjects.

TABLE II. Percentage of subjects responding different noise annoyance in two of the clusters in Fig. 5.

	Population	Not specific/FN	Not FN
not at all	39%	3%	94%
slightly	31%	11%	0%
fairly	17%	69%	0%
strongly	11%	17%	0%
extremely	2%	0%	6%
No. subjects	2472	379	890

ficity and no false negative is obtained only when predicting no annoyance at all or extreme annoyance. High false negative is more often found in combination with nonspecificity when predicting the middle noise annoyance categories. This result is not unexpected. Also for the human expert it is easier to be specific when predicting the extremes of the annoyance scale, while it is rather difficult to differentiate situations that can go either way.

In the next section, the FRB will be used to analyze the importance of hypotheses concerning noise annoyance, which are formulated as fuzzy rules. When FRBs including different sets of rules are to be compared, it is necessary to tune both models. The reason for this is that input data processed by the rules are not necessarily orthogonal. Assume, for example, that a FRB contains two rules operating on such nonorthogonal data that partly compensate. Tuning may include both rules to minimize model error. When one of the rules is then omitted this could lead to a very bad model which could be falsely interpreted as proof for the rule being very important. After tuning the reduced system, the compensating rule will also be removed, showing that both rules are probably not representing hypotheses that lead to a better prediction of noise annoyance.

It may sound instructive to consider the optimized sufficiency, λ, of each rule as a measure of importance of that rule. However since the λ are the variables in an optimization process, they are much more sensitive to numerical noise than the dependent variables, the measures of quality. Moreover, a rule may be sufficient but at the same time only apply to a very limited number of subjects in the data set, thus making it quite useless.

In conclusion, comparing different sets of hypotheses using the FRB is best done by comparing quality measures, after optimization.

VI. EXAMPLE RESULTS

A. The influence of adding additional information and knowledge

It is very popular to base the prediction of average noise annoyance in a population on a single noise level indicator, such as L_{dn}. Although it is clear that this indicator is insufficient to estimate annoyance perceived by a single individual or a small group, it is at least a good starting point for a comparison.

To evaluate the validity and importance of knowledge and corresponding input information that is added to the FRB, the increase in quality of the model is investigated

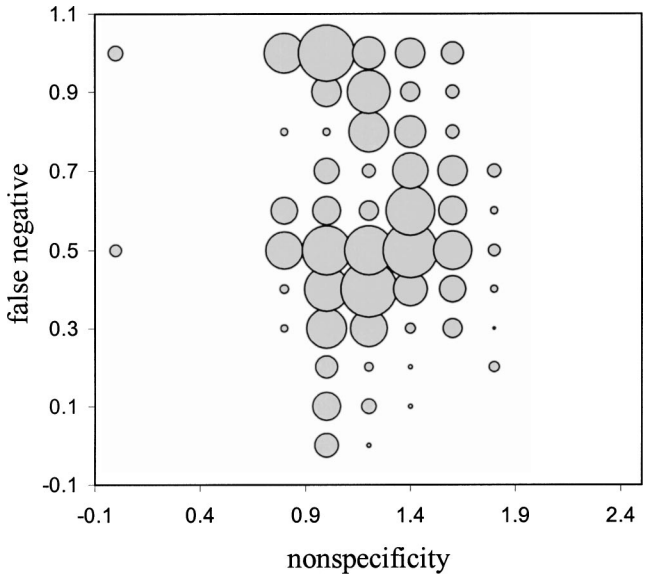


FIG. 7. Distribution of the subjects over a false negative versus nonspecificity plane for a model containing only rules based on L_{dn} that is tuned with $\alpha=0.25$; the area of the bubbles is proportional to the number of subjects.

(Flemish data set). In a purely fuzzy approach, quality means a combination of low nonspecificity and low false negative rating (Sec. V A). Because tuning may result in a more specific or a more accurate model depending on the focus α that is used while optimizing for a particular data set, the comparison of models has to be done in this two-dimensional plane. For a FRB containing only rules based on L_{dn}, that is tuned using $\alpha=0.25$, the population is distributed as shown in Fig. 7. When compared to Fig. 5, where the results obtained in Sec. V D for a FRB including many fuzzy rules, it is obvious that the rules based on L_{dn} alone do not allow obtaining both low nonspecificity and low false negative for the same fraction of subjects. Before drawing any conclusions on the quality of fuzzy rules (hypotheses on field knowledge), it must be verified how different tuning affects the performance of the model. By changing α, the prediction error, e, shown in Fig. 8 is obtained. The trend line connect-

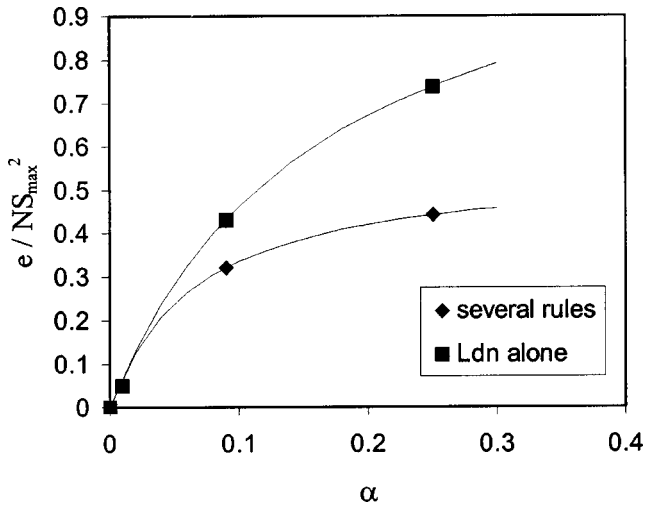


FIG. 8. Prediction error as a function of tuning parameter α for two FRB models.

ing the points in the figure has the form $\alpha/(C\alpha+D)$ since this is the trend that can be expected for single rule FRBs. For low α the prediction error for both models coincides. For these α, very nonspecific models that simply state that all labels are equally possible outcomes of asking a subject about noise annoyance are obtained. Hence it makes no difference which rules are included. As α increases, the superior performance of the model including additional rules becomes obvious. From this analysis, it can be concluded that the introduction of additional rules in this case leads to a model that scores better on both quality measures at the same time. Note, these conclusions can only be drawn for the importance of the rules and not for the information content of the variables that they use. Although the knowledge incorporated in fuzzy rules is usually very nonlinear and it is expressed using words, this does not guarantee that it is of high quality. However, it is safe to state that the combination of distance to the roads, degree of urbanization, reported traffic density, household size, age, and gender contain additional information on road traffic noise annoyance that a person reports in this social survey.

For easy comparison with classical models, the percentage of the subjects whose noise annoyance is correctly predicted is also included, although this is against the philosophy that annoyance is inherently a vague concept. To defuzzify H, a similarity measure is used. The result is weighted with $O(s)$ to take into account occurrence of different road traffic noise annoyance levels in the database. This results in 43% correct predictions by the model including all rules and 36% correct predictions for the model only containing L_{dn} based rules.

B. Noise sensitivity as an example of an internal variable

Sensitivity to noise is found to be significantly linked with noise annoyance in most noise surveys.[10,48–50] Particularly high noise sensitivity seems to be a relative stable personality trait over time.[48,50,51] It is therefore chosen as an example of a preconditioner that cannot be made external to the model but that is estimated internally on the basis of a number of well-chosen indicators that become model inputs themselves. For noise sensitivity, a two-step process can be used since reported noise sensitivity is available in many surveys. The first step consists of relating reported noise sensitivity to possible indicators using a number of rules (Table I). These rules are inspired by the work reported in Refs. 48, 52, 36, and 50. The second step is to include the outcome of the model for noise sensitivity in the FRB model for noise annoyance (Table I). In the second step, the linguistic approximation and interpretation of the fuzzy estimate A_{sens} of noise sensitivity is not required. A_{sens} is used directly as an input to subsequent rules automatically taking into account the uncertainty and vagueness in this result. In an extreme case, if there is insufficient evidence to draw conclusions on noise sensitivity, A_{sens} will be very nonspecific and will not trigger any subsequent fuzzy rules based on this variable. This is a great strength of the FRB.

The Inn Valley survey is used to tune and test this model since it contains a question asking explicitly for noise sensitivity, providing a verbal reply scale labeled ''überhaupt nicht,'' ''gering,'' ''mittelmässig,'' and ''stark'' (not at all, slightly, moderately, strongly). Adding the fuzzy rules, based on this reported sensitivity to a FRB that contains only L_{dn} based rules, increases performance. Expressed as percentage correctly predicted annoyance labels, road traffic noise annoyance is correctly predicted for 35.4% of the population (32.5% without sensitivity rules) while rail traffic noise annoyance is predicted correctly for 38.7% of the population (37.5% without sensitivity rules). These percentages are given as an indication only. The advantage of including noise sensitivity is subtler. It allows us to model ''difficult'' to understand differences in response in the in-between annoyance region.

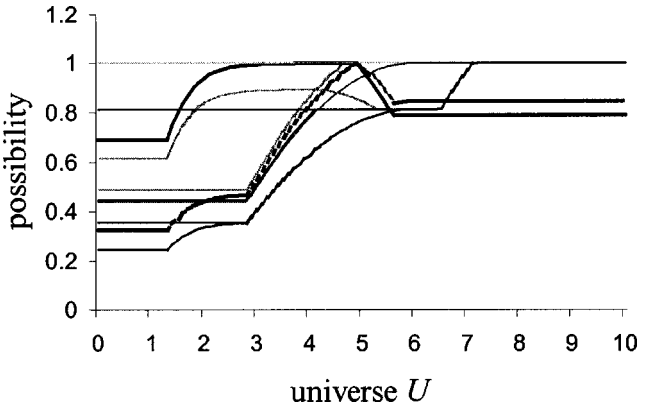


FIG. 9. Membership functions for A_{sens} for a random selection of subjects in the database.

Only very few indicators are available that allow estimating whether a person has a higher possibility of being sensitive to noise. The rules in Table I allow predicting the response to the noise sensitivity question in the Inn Valley survey for about 30% of the subjects. In Fig. 9 the membership function for A_{sens} is shown for a dozen subjects. This illustrates how very vague information on noise sensitivity—estimated on the basis of a limited number of indicators—is transferred inside the model. It may be impossible to classify all subjects in groups for which the combination of indicators results in a clear estimation of their noise sensitivity. The extent to which this information is available is automatically included in the inference based on fuzzy rules with A_{sens} in their antecedent. In the extreme case where A_{sens} is *unknown*, none of the fuzzy rules based on this variable are triggered.

C. How general is the fuzzy model?

The knowledge incorporated in the FRB proposed in this paper is not extracted from the data but proposed, as a kind of hypothesis by the expert, based on literature. Tuning involves only modifying the sufficiency of each rule to create the possibility for noise annoyance. Nevertheless, overfitting on a particular data set cannot be completely excluded. To find out exactly how general the FRB model is (after it is tuned to a particular data set), prediction of railway noise annoyance as obtained from the Flemish and Tirol surveys is used. Recorded data matches exactly between both studies for a few variables only. In order not to obscure results by

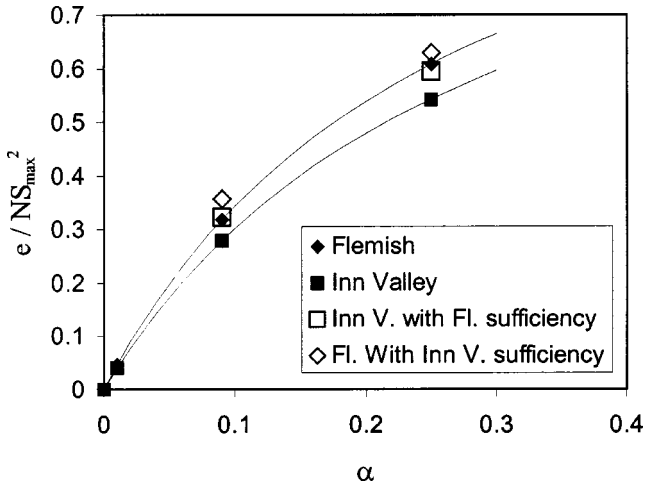


FIG. 10. Scaled prediction error as a function of parameter a for Flemish and Tirol rail traffic noise annoyance prediction; trend lines are for tuned models.

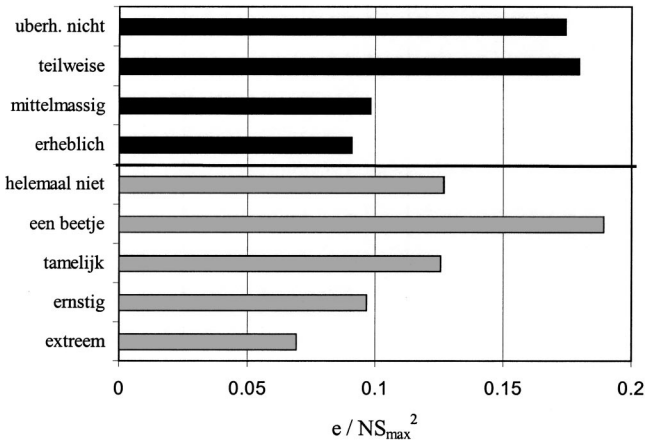


FIG. 11. Prediction error for different categories of reported noise annoyance in the Inn Valley and the Flemish surveys; category labels: überhaupt nicht (not at all), teilweise (partially), mittelmassig (moderately), erheblich (strongly), helemaal niet (not at all), een beetje (a little), tamelijk (fairly), ernstig (strongly), extreem (extremely); translation based on Ref. 12.

assumptions on input data, only L_{dn}, distance to the railroad, age, and gender were used as input variables to their corresponding rules. A particular difference between the two surveys is that the Flemish one used five Dutch labels while the Tirol one used four German labels as answer categories.

Tuning was done on the Flemish database and the model was subsequently run also on the Tirol data changing only the category labels. Because predicting response on a five-label scale is inherently more difficult than predicting the response on a four-label scale, the prediction error e is divided by the highest nonspecificity for each scale NS_5 and NS_4. The exercise is repeated for several values of the parameter α. Figure 10 shows the results. Prediction error for the Tirol data produced by a model tuned to Flemish data (open squares) is surprisingly similar to the prediction error obtained on the Flemish data (closed diamond). Optimization on the Tirol data results in lower prediction error (closed squares), but the rule sufficiency tuned to the Tirol data makes performance of the model on the Flemish data slightly worse (open diamond). A possible explanation could be that the Tirol data are all taken in the vicinity of the same major rail way track following the Inn Valley while the Flemish data include railway tracks with different type and density of rolling material on them. This makes the Tirol data more specific and thus the tuning slightly overfitted to this particular situation.

Figure 11 shows scaled prediction error, e, for each category of reported railway noise annoyance. The FRB model is tuned for the corresponding survey using $\alpha=0.25$, resulting in a *specific* model. The model clearly performs better in predicting high levels of annoyance as was expected. The high prediction error for the lowest levels was somewhat unexpected. Looking in detail at the simulations revealed that the model attributes higher possibility to annoyance levels that are higher than the reported level.

The FRB model seems to generalize quite well, at least on the data used here. The fuzziness in the outcome of the model and evaluation based on purely fuzzy quality measures allows the model to give a very uncertain result in those situations that are not absolutely clear. This is not possible using crisp noise annoyance prediction. Moreover, fuzzy relations between factors and annoyance are not optimized for a particular data set; they are merely included or excluded by tuning. Both features explain why the FRB model generalizes quite well.

D. Aggregation over a small population

Although the FRB model proposed in this work aims at predicting annoyance on an individual basis with the ultimate purpose of helping to understand the mechanisms involved, it can also be a tool for predicting the effect on ''small'' populations. To aggregate the prediction made for each individual in the population, a direct fuzzy extension of the statistical distribution of crisp categorical data is proposed. The percentage of the population is counted that has a *high* possibility, $\pi_p(i)$, for reporting annoyance level L_i, for all i. By using different threshold $\pi_{\min}$ for defining *high*, $\pi_p(i) > \pi_{\min}$, some fine tuning is possible. Note that each individual can contribute to percentage counts for several L_i at the same time. This fully takes into account the fact that noise annoyance may be completely unpredictable for some individuals in the (small) population.

As an example, the prediction by the FRB model using the same set of rules as in Sec. V D is aggregated over city centers, over the whole region, and over rural areas. The results, using $\pi_{\min}=0.5$, are shown in Fig. 12. The percentage of the population for which the possibility for reporting *no annoyance at all* is high, is much higher in the rural area than the Flemish average. In the city center, this percentage is clearly lower. *Extreme or strong* annoyance is a possible answer for a larger percentage of people in city centers than in rural areas.

Figure 12 also includes the percentage of the population that actually reported the indicated level of annoyance in the survey for each of the selections considered. In general the trend produced by the model is comparable with the survey data. As expected the FRB obtains a high possibility for a higher percentage of the population than the actual report.

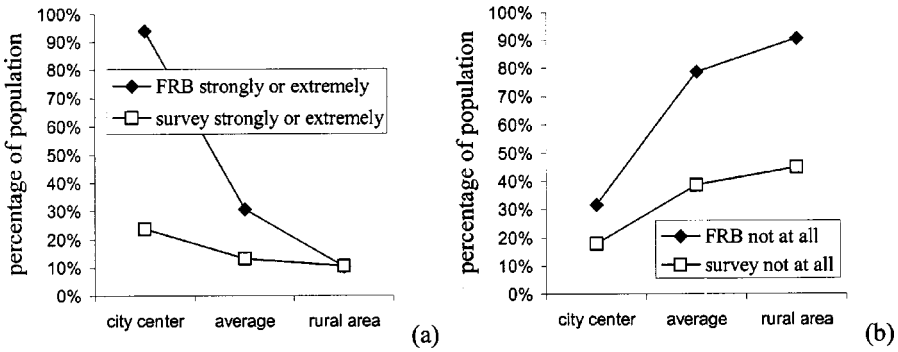


FIG. 12. Aggregation of the FRB model prediction over the city centers, the whole region (average), and urban communities, compared to the results of the survey (Flanders); (a) extreme or strong road noise annoyance; (b) no road noise annoyance at all.

This indicates that the model cannot actually pinpoint the individuals within the area. The ratio of percentage possibly annoyed (model) to the percentage actually annoyed (survey) is a measure of how well the FRB model can locate subjects. For the category ''not at all annoyed,'' this ratio is lower because this condition can be attributed to the part of the population not at all exposed to noise, with great certainty. Extreme or strong annoyance is only reported by part of the population exposed to high noise levels (about one in three). Additional nonexposure rules have to be used to discriminate within this group, which is obviously a more difficult task for the model. Hence there is less certainty and a higher percentage of the population is labeled possibly extremely or strongly annoyed.

VII. CONCLUSIONS

Annoyance caused by traffic noise is inherently a vague concept. This was used as a starting argument for this work. Based on the mathematical theory of fuzzy sets and fuzzy rule based systems a framework was proposed to model noise annoyance. Domain knowledge, even if not surpassing the stage of a hypothesis, was incorporated in the model in the form of rules that can be formulated using natural language. This rule based model was made tunable by adding sufficiency to each of the rules in the system. This tuning as well as comparison to survey data required a whole new approach to the quality of prediction that conformed to the premise that noise annoyance is inherently vague. A fuzzy extension of *false negative* combined with the requirement that the model must be minimally decisive were shown to be useful measures of quality.

A few examples illustrated the application of the new modeling framework. It was demonstrated that the model could be used to estimate the importance of additional soundscape descriptors and additional internal or external nonacoustic factors for predicting the reaction of individuals to traffic noise. These variables can even be more complex constructs based on indirect indicators (e.g., sensitivity to noise). Furthermore, the utility for comparing surveys formulated in another language or using a different categorical answering scale was demonstrated. By getting accurate predictions from this model for two different surveys in different regions with strikingly diverging geographical features, it was proven that inherent vagueness and uncertainty were accurately modeled and no unfounded prediction claims were made.

Eventually, it was also demonstrated that with the use of fuzzy set mathematics the model results could be further processed to obtain derived results (e.g., the noise impact on small populations such as a selected community). Combined with exposure calculations that fully account for uncertainty,[53] this approach could lead to a new type of environmental impact assessment and noise policy making support that incorporate vagueness and uncertainty in the decision process.

[1] H. S. Koelega (Ed.), *Environmental Annoyance: Characterization, Measurement, and Control. Amsterdam* (Elsevier, The Netherlands, 1987).
[2] H. M. E. Miedema and H. Vos, ''Exposure response functions for transportation noise,'' J. Acoust. Soc. Am. **104**, 3432–3445 (1998).
[3] T. J. Schultz, ''Synthesis of social surveys on noise annoyance,'' J. Acoust. Soc. Am. **64**, 377–405 (1978).
[4] S. Fidell, D. S. Barber, and T. J. Schultz, ''Updating a dosage-effect relationship for the prevalence of annoyance due to general transportation noise,'' J. Acoust. Soc. Am. **89**, 221–233 (1991).
[5] I. D. Griffiths and G. J. Raw, ''Adaptation to changes in traffic noise exposure,'' J. Sound Vib. **132**, 331–336 (1989).
[6] I. D. Griffiths and G. J. Raw, ''Community and individual response to changes in traffic noise exposure,'' J. Sound Vib. **111**, 209–217 (1986).
[7] G. J. Raw and I. D. Griffiths, ''Subjective response to changes in road traffic noise: a model,'' J. Sound Vib. **141**, 43–54 (1990).
[8] R. Klæboe, M. Kolbenstvedt, P. Lercher, and S. Solberg, ''Changes in noise reactions—evidence for an area-effect?'' in *Proceedings of Internoise 98*, Christchurch, New Zealand (1998).
[9] T. Gjestland, ''Regional differences in noise annoyance assessments,'' in *Proceedings of Internoise 98*, Christchurch, New Zealand (1998).
[10] J. M. Fields, ''Effect of personal and situational variables on noise annoyance in residential areas,'' J. Acoust. Soc. Am. **93**, 2753–2763 (1993).
[11] P. Lercher, ''Deviant dose-response curves for traffic noise in 'sensitive areas'?'' in *Proceedings of Internoise 98*, Christchurch, New Zealand (1998).
[12] D. Botteldooren, A. Verkeyn, C. Cornelis, and M. De Cock, ''On the meaning of noise annoyance modifiers: a fuzzy set theoretical approach,'' Acust. Acta Acust. **88**(2), 239–251 (2002).
[13] J. M. Fields, R. G. De Jong, T. Gjestland, I. H. Flindell, R. F. S. Job, S. Kurra, P. Lercher, M. Vallet, T. Yano, R. Guski, U. Felscher-Suhr, and R. Schumer, ''Standardized general-purpose noise reaction questions for community noise surveys: research and a recommendation,'' J. Sound Vib. **242**(4), 641–679 (2001).
[14] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications* (Prentice-Hall, Upper Saddle River, NJ, 1995).
[15] E. Tsiporkova and H.-J. Zimmermann, ''Aggregation of compatibility and equality: a new class of similarity measures for fuzzy sets,'' in *Proceedings of IPMU-Information Processing and Management of Uncertainty in Knowledge-based Systems Seventh International Conference*, Paris (1998), pp. 1769–1776.
[16] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems* (Prentice-Hall, Upper Saddle River, NJ, 2001).
[17] Modus Ponens (*Latin*: mode that affirms). A classical logic argument form: If P, Then Q. P. Therefore, Q.
[18] D. Dubois and H. Prade, ''Fuzzy sets in approximate reasoning, Part I:

Inference with possibility distributions,'' Fuzzy Sets Syst. **40**, 143–202 (1991).
[19] D. Dubois and H. Prade, ''What are fuzzy rules and how to use them,'' Fuzzy Sets Syst. **84**, 169–185 (1996).
[20] L. Zadeh, ''Pruf—a meaning representation language,'' in *Fuzzy Reasoning and its Applications*, edited by E. Mamdani and R. Gaines, Computer and People Series (Academic, London, 1981), pp. 1–66.
[21] R. Klaeboe, ''The possible effect of the neighbourhood soundscape on exposure-effect relationships,'' in *Proceedings of Inter Noise 2001*, Den Haag, The Netherlands (2001), Vol. 3, pp. 1499–1504.
[22] B. Berglund, P. Hassmén, and A. Preis, ''Annoyance and spectral contrast are cues for similarity and preference of environmental sounds,'' J. Sound Vib. **250**, 53–64 (2002).
[23] A. Fyhri and R. Klaeboe, ''Exploring the impact of visual aeshetics on the soundscape,'' in *Proceedings of Internoise 99*, Fort Lauderdale, FL (1999).
[24] S. Viollon, C. Lavandier, and C. Drake, ''Influence of visual setting on sound ratings in an urban environment,'' Appl. Acoust. **63**, 493–511 (2002).
[25] J. Carles, F. Bernaldez, and J. De Lucio, ''Audio-visual interactions and soundscape preferences,'' Landscape Res. **17**, 52–56 (1992).
[26] M. Haider, E. Kundi, E. Groll-Knapp, and M. Koller, ''Interactions between noise and air pollution,'' Environ. Int. **16**, 593–601 (1990).
[27] P. Lercher, R. Schmitzberger, and W. Kofler, ''Perceived traffic air pollution, associated behaviour and health in an alpine area,'' Sci. Total Environ. **169**, 71–74 (1995).
[28] R. Klæboe, M. Kolbenstvedt, J. Clench-Aas, and A. Bartonova, ''Oslo traffic study—part 1: an integrated approach to assess the combined effects of noise and air pollution on annoyance,'' Atmos. Environ. **34**, 4727–4736 (2000).
[29] K. Zeichart, ''Kombinatorische Wirkungen von Bahnlärm und Bahnerschütterungen,'' (''Combined effect of traffic noise and vibrations''), Z. Lärmbekämpfung **45**, 7–16 (1998).
[30] R. Paulsen and J. Kastka, ''Effects of combined noise and vibration on annoyance,'' J. Sound Vib. **181**, 295–314 (1995).
[31] E. Öhrström, ''Effects of exposure to railway noise—A comparison between areas with and without vibration,'' J. Sound Vib. **205**, 555–560 (1997).
[32] W. Passchier-Vermeer, ''Vibrations in the environment. Factors related to vibration perception and annoyance,'' TNO-Report 98.022. Leiden: TNO (1998).
[33] H. M. E. Miedema and H. Vos, ''Demographic and attitudinal factors that modify annoyance from transportation noise,'' J. Acoust. Soc. Am. **105**, 3336–3344 (1999).
[34] R. S. F. Job, ''Community response to noise: A review of factors influencing the relationship between noise exposure and reaction,'' J. Acoust. Soc. Am. **83**, 991–1000 (1988).
[35] R. Guski, ''Personal and social variables as co-determinants of noise annoyance,'' Noise Health **3**, 45–56 (1999).
[36] P. Lercher, ''Environmental noise and health: An integrated research perspective,'' Environ. Int. **22**, 117–129 (1996).
[37] S. L. Staples, ''Human response to environmental noise,'' Am. Psychol. **51**, 143–150 (1996).
[38] R. F. S. Job, J. Hatfield, N. L. Carter, P. Peploe, R. Taylor, and S. Morrell, ''Reaction to Noise: The Roles of Soundscape, Enviroscape and Psychscape,'' in *Proceedings of Internoise 99*, Fort Lauderdale, FL, 6–8 December 1999, pp. 1309–1314.
[39] B. Schulte-Fortkamp and W. Nitsch. ''On soundscapes and their meaning regarding noise annoyance measurements,'' in *Proceedings of Internoise 99*, Fort Lauderdale, FL, 6–8 December 1999.
[40] J. M. Fields and F. L. Hall, ''Community effects of noise,'' in *Transportation noise, Chapter 3*, edited by P. M. Nelson (Butterworth, London, 1986).
[41] P. Lercher and W. Kofler, ''Behavioral and health responses associated with road traffic noise exposure along alpine through-traffic routes,'' Sci. Total Environ. **189/190**, 85–89 (1996).
[42] P. Lercher, ''Context and coping as moderators of potential health effects in noise-exposed persons,'' in *Advances in Noise Series. Vol. I: Biological Effects*, edited by D. Prasher and L. Luxon (Whurr, London, 1998), pp. 328–335.
[43] I. Van Kamp, ''Coping with noise and its consequences,'' Ph.D. Dissertation, Groningen, Styx & PP Publications, 1990.
[44] W. Bandler and L. Kohout, ''Fuzzy power sets and fuzzy implication operators,'' Fuzzy Sets Syst. **4**, 13–30 (1980).
[45] M. Higashi and G. J. Klir, ''Measures of uncertainty and information based on possibility distributions,'' Int. J. Gen. Syst. **9**(1), 43–58 (1983).
[46] B. Fogel, *Evolutionary Computing: Toward a New Philosophy of Machine Intelligence*, 2nd ed. (IEEE, New York, 2000).
[47] A. Verkeyn, D. Botteldooren, G. De Tre, and R. De Caluwe, ''Fuzzy modeling of traffic noise annoyance,'' in *Proceedings of the Joint 9th IFSA World Congress and the 20th NAFIPS International Conference*, Vancouver, Canada, July 2001, pp. 1176–1181.
[48] S. A. Stansfeld, ''Noise, noise sensitivity and psychiatric disorder: epidemiological and psychophysiological studies,'' in *Psychological Medicine Monograph Supplement 22* (Cambridge U.P., Cambridge, 1992).
[49] D. Halpern, *Mental Health and the Built Environment. More than Bricks and Mortar*? (Taylor & Francis, London, 1995).
[50] S. A. Stansfeld, M. M. Haines, M. Burr, B. Berry, and P. Lercher, ''A review of environmental noise and mental health,'' Noise Health **8**, 55–58 (2000).
[51] S. A. Stansfeld, ''Noise sensitivity, depressive illness and personality: a longitudinal study of depressed patients with matched control subjects,'' in *Noise as a Public Health Problem*, edited by B. Berglund, U. Berglund, J. Karlsson, and T. Lindvall (Swedish Council for Building Research, Stockholm, Sweden, 1988), Vol. 3, pp. 339–344.
[52] M. E. Nivison and I. M. Endresen, ''An analysis of relationships among environmental noise, annoyance and sensitivity to noise, and the consequences for health and sleep,'' J. Behav. Med. **16**, 257–276 (1993).
[53] T. De Muer and D. Botteldooren, ''Fuzzy noise maps,'' in *Proceedings of Forum Acusticum*, Sevilla, Spain (2002).

Cochlear compression estimates from measurements of distortion-product otoacoustic emissions

Stephen T. Neely,[a] Michael P. Gorga, and Patricia A. Dorn
Boys Town National Research Hospital, Omaha, Nebraska 68131

(Received 19 December 2002; revised 30 June 2003; accepted 2 July 2003)

Evidence of the compressive growth of basilar-membrane displacement can be seen in distortion-product otoacoustic emission (DPOAE) levels measured as a function of stimulus level. When the levels of the two stimulus tones (f_1 and f_2) are related by the formula $L_1 = 39$ dB $+ 0.4 \cdot L_2$ [Kummer *et al.*, J. Acoust. Soc. Am. **103**, 3431–3444 (1998)] the shape of the function relating DPOAE level to L_2 is similar (up to an L_2 of 70 dB SPL) to the classic Fletcher and Munson [J. Acoust. Soc. Am. **9**, 1–10 (1933)] loudness function when plotted on a logarithmic scale. Explicit estimates of compression have been derived based on recent DPOAE measurements from the laboratory. If DPOAE growth rate is defined as the slope of the DPOAE I/O function (in dB/dB), then a cogent definition of *compression* is the reciprocal of the growth rate. In humans with normal hearing, compression varies from about 1 at threshold to about 4 at 70 dB SPL. With hearing loss, compression is still about 1 at threshold, but grows more slowly above threshold. Median DPOAE I/O data from ears with normal hearing, mild loss, and moderate loss are each well fit by log functions. When the I/O function is logarithmic, then the corresponding compression is a linear function of stimulus level. Evidence of cochlear compression also exists in DPOAE suppression tuning curves, which indicate the level of a third stimulus tone (f_3) that reduces DPOAE level by 3 dB. All three stimulus tones generate compressive growth within the cochlea; however, only the relative compression (RC) of the primary and suppressor responses is observable in DPOAE suppression data. An RC value of 1 indicates that the cochlear responses to the primary and suppressor components grow at the same rate. In normal ears, RC rises to 4, when f_3 is an octave below f_2. The similarities between DPOAE and loudness compression estimates suggest the possibility of predicting loudness growth from DPOAEs; however, intersubject variability makes such predictions difficult at this time.
[DOI: 10.1121/1.1604122]

PACS numbers: 43.64.Ha, 43.64.Jb [BLM]

I. INTRODUCTION

Under normal conditions, the mechanical responses of the mammalian cochlea are highly nonlinear. At low stimulus levels (i.e., at levels near threshold), the response grows approximately linearly with level, but becomes compressive as level increases (see Robles and Ruggero, 2001, for a comprehensive review of cochlear mechanical responses). These observations are based on data collected invasively from lower animals. It is known, however, that the human cochlea also behaves nonlinearly. There are several lines of evidence in support of this view, including observations based on behavioral masking data (Oxenham and Plack, 1997, 1998, 2000; Plack and Oxenham, 1998; Nelson *et al.*, 2001) and measures of growth of loudness (Buss *et al.*, 1998).

Evidence supporting nonlinear cochlear function in the normal human cochlea also comes from measurements of otoacoustic emissions (OAEs). OAEs are low-level, acoustic responses that can be measured in the ear canals of humans with normal cochlear function. They are evidently produced by the outer hair cell (OHC) system (Brownell, 1990) and possess characteristically nonlinear properties. These responses are reduced or absent when OHC damage exists. Most insults that cause damage to the cochlea affect the OHCs first, and OHC damage is always accompanied by threshold elevation (i.e., hearing loss). In humans, it is never possible to directly assess cochlear status *in vivo*. However, the close linkage between OHC damage and hearing loss, coupled with the close linkage between the production of OAEs and OHC status, suggests that measurements of OAEs might provide a window into nonlinear cochlear behavior in humans.

Indeed, a considerable literature has developed that demonstrates the relationship between OAEs and hearing loss in humans (see Harris and Probst, 2002 and Gorga *et al.*, 2002, for reviews). In the presence of normal hearing, humans, like other mammals, produce robust OAEs, indicating that nonlinear behavior is normal and, therefore, OHC function is normal. In humans with hearing loss, OAE levels are reduced or the response is absent, indicating that nonlinear behaviors, present in the normal condition, are absent when hearing loss exists. As a consequence of these observations, OAE measurements have been used to determine auditory status dichotomously as normal or impaired. This application has been particularly useful in newborn hearing screening (e.g., Norton *et al.*, 2000a, b), where the goal is to determine which infants may have hearing loss.

There also have been efforts to go beyond determining

[a]Electronic mail: neely@boystown.org

the likelihood that hearing loss exists, and to determine the magnitude of the hearing loss, using OAE measurements. These efforts have included (1) comparisons between distortion product otoacoustic emission (DPOAE) thresholds and pure-tone thresholds (Martin *et al.*, 1990; Stover *et al.*, 1996; Dorn *et al.*, 2001); (2) correlations between DPOAE level for moderate-level stimuli and pure-tone threshold (Martin *et al.*, 1990; Gorga *et al.*, 2002); (3) multivariate approaches to predicting thresholds (Kimberley *et al.*, 1994, 1997); and (4) examination of the entire DPOAE input/output (I/O) function to predict threshold (Boege and Janssen, 2002; Oswald *et al.*, 2002; Gorga *et al.*, 2003a). Each of these approaches has resulted in various degrees of success, in spite of the variability associated with the measurements, suggesting that the level of DPOAE responses is graded according to cochlear status.

Recently, psychoacoustic data have been described in which it appears that a comparison of the masking observed for on-frequency and low-frequency maskers can be used to provide a measurement of peripheral nonlinearities in humans (Plack and Oxenham, 1998; Oxenham and Plack, 1997, 2000; Nelson *et al.*, 2001). Underlying this approach is the view that when cochlear function is normal, the response of the cochlea is compressed (nonlinear) when a particular place is driven at its characteristic frequency, while the response at that same place is more linear when it is driven by a frequency much lower than its characteristic frequency. Thus, an examination of the masking effect from a low-frequency masker (relative to the characteristic frequency) on the threshold to a signal (which is represented at the characteristic place) provides an estimate of the nonlinear response. Psychoacoustic measurements can be time consuming and typically require a level of cooperation that many subjects, such as infants and young children, cannot provide.

A similar approach has been applied to DPOAEs. A DPOAE suppression tuning curve (STC) is defined as the suppressor level that decreases the DPOAE level by a criterion amount as a function of suppressor frequency. Mills (1998) and Pienkowski and Kunov (2001) measured DPOAE STCs, and compared the suppressor levels on the tail to those at the tip of the STC to obtain tip-to-tail differences. These tip-to-tail differences were viewed as estimates of cochlear amplifier ''gain'' and were in the range of values that would be predicted from theoretical estimates of cochlear-amplifier gain (e.g., Neely and Kim, 1986). They are also within the range of values from more direct studies in lower animals (e.g., Ruggero and Rich, 1991).

We have also used DPOAE measurements to estimate the compressive characteristics of the human cochlea. In one study, DPOAE I/O functions were measured over a wide range of levels in subjects with normal hearing and subjects with hearing loss (Dorn *et al.*, 2001). DPOAE I/O functions in normal ears were similar in shape to basilar-membrane mechanical responses from lower animals (Ruggero and Rich, 1991; Ruggero *et al.*, 1997). That is, they were characterized by a low-level linear portion, followed by a compressive region for moderate stimulus levels. Subjects with hearing loss produced I/O functions that were characterized by threshold elevation, a reduced amount of compression for moderate levels, and reduced range of input levels over which compression took place.

Additionally, DPOAE STC measurements were used to describe the tip-to-tail difference in subjects with normal (Gorga *et al.*, 2002) and impaired (Gorga *et al.*, 2003b) hearing. This approach was based on observations made by Mills (1998) and Pienkowski and Kunov (2001), but extended their work by including a wider range of levels and by making these measurements in both normal and impaired human ears. Compression was evident in these measurements, in that the tip-to-tail difference decreased as level increased (Gorga *et al.*, 2002, 2003b), an effect that was particularly apparent in ears with normal hearing. The sharpness of STC tuning was quantified in terms of Q_{10} and Q_{ERB}. The Q_{10} of a filter is defined as its center frequency divided by its 10-dB bandwidth. Likewise, the Q_{ERB} of a filter is its center frequency divided by its equivalent-rectangular bandwidth. Interestingly, estimates of Q_{10} and Q_{ERB} were similar in impaired and normal ears at those stimulus levels for which comparisons could be made. If anything, there appeared to be a slight decrease in the tip-to-tail difference among ears with hearing loss, along with a corresponding slight increase in Q_{10} and Q_{ERB}. This suggests that the responses to a low-frequency and an on-frequency suppressor were more similar in ears with hearing loss. Since cochlear insults have a larger effect on the response to on-frequency stimuli, compared to low-frequency stimuli (see Kiang *et al.*, 1976; Dallos and Harris, 1978; Liberman and Dodds, 1984; Ruggero and Rich, 1991, or any of the many other papers that describe basilar-membrane or single-unit responses in the presence of OHC damage), then this effect must be the result of an increase in response growth (a decrease in compression) for the on-frequency condition.

A more complete understanding of nonlinear cochlear processing in humans might lead to insights related to factors that influence perceptual effects, such as loudness. Clearly, humans are capable of encoding a wide range of levels; it is likely that compression of moderate level stimuli is an important factor contributing to this wide dynamic range (e.g., Allen and Neely, 1992). Furthermore, it is known that some patients with hearing loss have unusually small dynamic ranges, resulting in abnormal percepts of loudness. Amplification systems that include compression attempt to compensate for this reduced dynamic range. However, mapping the characteristics of hearing aids to the compressive needs of individual patients can be problematic, especially in infants and young children, for whom it may be difficult to articulate their percepts of loudness (e.g., Stelmachowicz *et al.*, 1998). For these patients, an objective approach that quantified the amount of compression in both normal-hearing and hearing-impaired humans could have important applications in the selection of hearing-aid characteristics.

The purpose of the present study was to examine the possibility of using DPOAE measurements to provide an estimate of peripheral compression in humans having either normal hearing or hearing loss. This paper represents analyses and modeling of recent DPOAE data that have been previously described (Dorn *et al.*, 2001; Gorga *et al.*, 2002). First, we characterize compression using a functional fit to

the data. Then, we compare our compression estimates to other estimates based on perceptual measures, such as the growth of loudness (Fletcher and Munson, 1933). We evaluate data from subjects with hearing loss to determine the extent to which the expected loss of compression is evident in DPOAE measurements. Finally, we apply our compression estimation methods to the measurements from individual subjects.

II. METHODS

A. Subjects

DPOAE I/O functions and STC tip-to-tail differences as a function of primary level are the two measurements that we use to develop estimates of compression. DPOAE I/O functions were measured in 27 normal-hearing and 50 hearing-impaired subjects (Dorn *et al.*, 2001), while DPOAE STC data were available on 13 subjects with normal hearing (Gorga *et al.*, 2002). For both studies, normal hearing was defined as audiometric thresholds of 20 dB HL or better (ANSI, 1996) for octave and half-octave frequencies from 0.25 to 8 kHz. Hearing loss was defined as thresholds greater than 20 dB HL at 4 kHz, which represented the f_2 frequency for both DPOAE I/O and STC measurements. Although there are indications that DPOAE level at a given frequency is influenced by hearing status at higher frequencies (e.g., Arnold *et al.*, 1999; Dorn *et al.*, 1999), this factor was not considered when assigning ears to hearing-loss categories. For both normal and impaired ears, middle-ear function was normal each day on which DPOAE measurements were made, based upon tympanometric evaluations.

B. Stimuli and DPOAE responses

Stimuli were generated using a high-quality soundcard (CardDeluxe, Digital Audio Labs) with 24-bit precision and a 32-kHz sampling rate. All signals were transduced by loudspeakers housed in a probe–microphone system (Etymotic, ER-10C), which also includes a microphone that is used during calibration and for recording the ear-canal DPOAEs. To enable the generation of higher-level stimuli, 20 dB of attenuation was removed from the output side of the probe–microphone system.

Responses were recorded by means of synchronous averaging using the same soundcard that produced the stimuli. Measurement-based stopping rules were used for all DPOAE measurements, such that data collection was terminated for a condition when the noise level was either below -25 dB SPL (for L_2 levels of 65 dB SPL or less) or when it was equal to the level of system distortion (for higher L_2 conditions). In either case, this stopping rule was designed to provide reliable data over the widest dynamic range possible without running the risk of misinterpreting system distortion for biological distortion. In addition, data collection stopped after 32 s of artifact-free averaging, even if the noise floor criterion was not met. This rule was needed so that averaging did not continue indefinitely because of high noise levels for any one condition. A series of cavity measurements, validated against measurements in patients with cochlear implants, was used to determine the level at which system distortion occurred. For more details related to estimates of system distortion, please see Dorn *et al.* (2001).

The DPOAE I/O stimulus consisted of two tones at frequencies f_1 and f_2. The DPOAE level was measured at the frequency $2f_1-f_2$. DPOAE I/O functions were measured for $f_2=4$ kHz with $f_1 \approx f_2/1.22$. The level of f_2 (L_2) ranged from -5 to 95 dB SPL. The level of f_1 (L_1) was set according to the equation $L_1=0.4\,L_2+39$ dB (Kummer *et al.*, 1998) for L_2 levels up to 65 dB SPL, beyond which $L_1=L_2$. The shape of the I/O function depends on the formula used to determine L_1. The Kummer *et al.* formula was chosen because, on average, it gives the largest DPOAE level in ears with normal hearing. Please see Dorn *et al.* (2001) for further details regarding the DPOAE I/O function measurements.

DPOAE STCs were derived from decrements in DPOAE level (amount of suppression) as a function of suppressor level. Sixteen suppressor frequencies (f_3) were used, while the probe (f_2) was held constant at 4 kHz. The level of the probe (L_2) was varied from 20 to 60 dB SPL in 10-dB steps, with L_1 set according to the equation $L_1=0.4\,L_2+39$ dB. The level of the suppressor (L_3) ranged from 5 to 85 dB SPL (5-dB steps). For additional details, please see Gorga *et al.* (2002, 2003b).

III. RESULTS AND DISCUSSION

Although we use DPOAE group-median data to illustrate our compression estimates, it is expected that the same methods can be applied to DPOAE measurements from individual ears. We will see, however, that variability in DPOAE measurements across ears, even within the same hearing-loss category, limits our ability to obtain compression estimates for individual ears.

A. DPOAE I/O functions in normal and impaired ears

The top panel of Fig. 1 shows median DPOAE level as a function of L_2 when $f_2=4$ kHz. This frequency was selected because it provided the widest dynamic range for DPOAE measurements and it was a frequency for which it was possible to find a sufficient number of subjects with hearing loss. The heaviest line represents median data from subjects with normal hearing, while lines of decreasing thickness represent data from subjects with thresholds of 25–35 and 40–50 dB HL at 4 kHz. These data are taken from Dorn *et al.* (2001). Note that the function for normal-hearing ears is characterized by a low DPOAE threshold (the stimulus level at which the DPOAE was first detectable above the noise floor), a relatively steep slope near threshold, a progressively more shallow slope as level increases up to about 80 dB SPL, followed by a steep slope for higher levels. In ears with hearing loss, there is an increase in DPOAE threshold, a steeper slope above threshold (compared to the normal I/O function), and a slope similar to that observed in normal ears for levels of 85 dB SPL and higher. The range of levels over which compressive growth is evident is reduced in the ears with hearing loss compared to the range observed in normal-hearing ears.

These median DPOAE I/O functions were fit with the following equation (Dorn *et al.*, 2001):

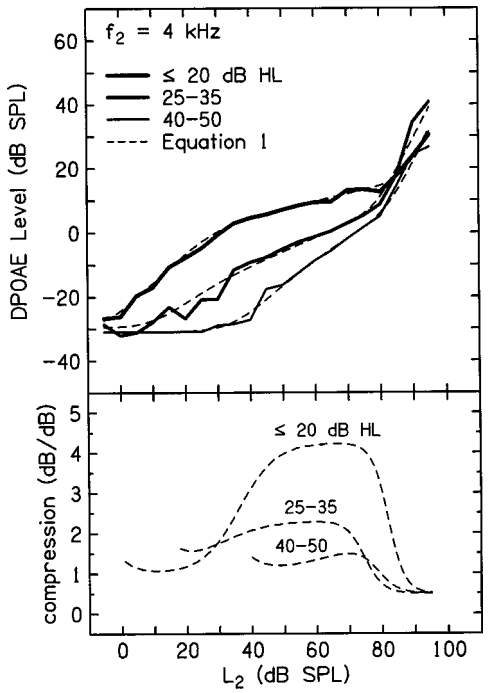


FIG. 1. DPOAE I/O functions for three hearing loss groups. The solid lines in the upper panel are median data based on measurements in ears with 20 dB HL or less, 25 to 35 dB HL, and 40 to 50 dB HL. The dashed lines in the top panel represent fits to the data using Eq. (1) in the text. The dashed lines in the lower panel show the compression corresponding to the fitted lines in the upper panel.

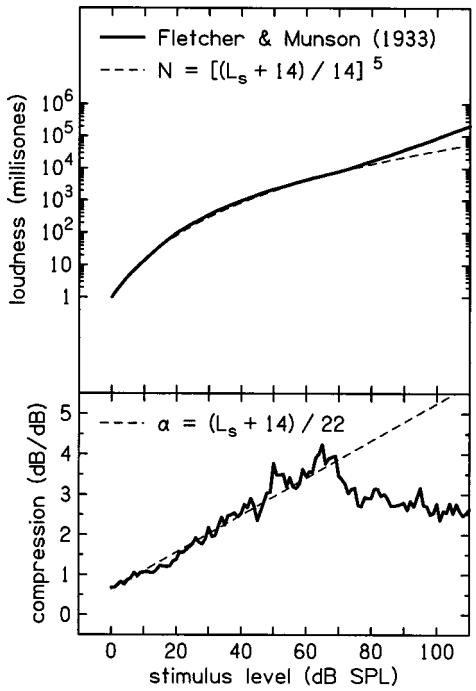


FIG. 2. Loudness as function of stimulus level. The solid line in the upper panel is based on loudness data presented in Fletcher and Munson (1933). The solid line in the lower panel is based on numerical differences between the loudness values at adjacent levels (1-dB steps). The dashed lines represent a fit to the loudness compression.

$$d(s)=n+\frac{g_0}{1+(g_1 s)^{1-p}}+(g_2 s)^2. \tag{1}$$

The variables s and $d(s)$ in this equation are related to the primary level L_2 and the $2f_1-f_2$ distortion-product level L_d by $L_2=10\cdot\log_{10}(s)$ and $L_d=10\cdot\log_{10}(d)$. The parameters n, p, g_0, g_1, and g_2 are determined by fitting the DPOAE I/O data. The parameter n can be interpreted as the background noise. The dashed lines in the top panel of Fig. 1 show examples of fitting Eq. (1) to the data from both normal and impaired ears. The lower panel shows the reciprocal of the slope of the fitted curves in the upper panel, which we call *compression* because it describes the dB change in the input level per dB change in the output level. Mathematically, we can define compression α_d for the DPOAE data by the equation

$$\alpha_d=\left(\frac{dL_d}{dL_2}\right)^{-1}. \tag{2}$$

Note that normal compression varies from about 1 at low levels (between 0 and 20 dB SPL) to a maximum of about 4 at moderate levels (between 50 and 70 dB SPL) in subjects with normal hearing. In ears with hearing impairment, the compressive region is restricted to a smaller range of input levels and its maximum is less than 4. In subjects with thresholds of 40–50-dB HL at 4 kHz, there is little evidence of compression.

B. Comparison of DPOAE growth to loudness growth

The normal DPOAE I/O function shown in the top panel of Fig. 1 is similar in some ways to the loudness data obtained by Fletcher and Munson (1933). The loudness data were collected at a different frequency (1 kHz) than the present measurements; however, there is no reason to believe that loudness or DPOAE growth differs substantially between 1 and 4 kHz. The upper panel of Fig. 2 plots the Fletcher–Munson loudness function for a 1-kHz tone (solid line) when hearing is normal. Note that there is a near-linear growth near threshold, followed by a compressive region for moderate-level stimuli. That is, loudness grows more slowly for moderate-level stimuli, much like the DPOAE I/O functions at the same levels. The reciprocal of the slope of the loudness function is plotted in the lower panel (also as a solid line). In this case, the compression curve is not smooth because it was numerically computed from the loudness values, which are in 1-dB steps. The amount of compression, according to this estimate, is similar to what was observed in the DPOAE data. Both achieve a maximum amount of compression at about 60 dB SPL, and in both cases, we would estimate the compression to be about 4 at this stimulus level. The two data sets differ for lower-level stimuli, in that the compression estimate from the loudness data increases gradually over a 60-dB range, compared to the compression estimate from the DPOAE I/O data, which makes a similar transition over about a 30-dB range. The two data sets also differ for higher-level stimuli, where the compression estimate from the loudness data remains relatively constant, while the compression evident in the DPOAE I/O data decreases as level increases above 70 dB SPL. The reasons for

this difference are not apparent, but might relate to issues associated with DPOAE measurements at such high levels. For example, it is unclear what high-level DPOAEs reflect, given that the outer hair cells (the source of the compression) function over a range of only about 60 dB. Furthermore, although special efforts were taken to assure that system distortion did not contaminate the DPOAE measurements (Dorn *et al.*, 2001), it is possible that the measurements were influenced by distortion that was not cochlear in origin.

If N is the perceptual loudness (e.g., in millisones) and L_s is the stimulus level, then we can define compression α_ℓ for the loudness data by the equation

$$\alpha_\ell = \left(\frac{dL_\ell}{dL_s}\right)^{-1}, \tag{3}$$

where $L_\ell = 10 \cdot \log_{10}(N)$ is a hypothetical *internal* sensation level. With this definition, compression of the loudness function (shown by the solid line in the lower panel of Fig. 2) varies from about 1 at low levels to a maximum of about 4 at a moderate level, which is similar to the compression we observed for the DPOAE I/O function. Note that loudness compression increases gradually with level up to about 70 dB SPL.

The loudness compression data follow nearly a straight line between stimulus levels of 0 and 70 dB SPL. We will take advantage of this feature of the data to develop the following linear approximation of loudness compression:

$$\alpha_\ell \approx (L_s + 14)/22. \tag{4}$$

This linear fit to compression suggests the following logarithmic fit to the loudness level:

$$L_\ell \approx 22 \cdot \ln((L_s + 14)/14). \tag{5}$$

Equation (5) includes an arbitrary constant that was selected to make $L_\ell = 0$ when $L_s = 0$. Equation (5) can be transformed using the definition of L_ℓ into the following equation that relates loudness N to signal level L_s:

$$N \approx ((L_s + 14)/14)^5. \tag{6}$$

The fit to the loudness data described by Eq. (6) is shown as a dashed line in the upper panel of Fig. 2 and appears to be a good approximation to the data for stimulus levels between 0 and 70 dB SPL. The numerical values in Eqs. (4)–(6) were selected on the basis of visual comparison with the data. In Eq. (6), we have arrived at a relatively simple relation between loudness and signal level by taking advantage of our observation that the measured loudness compression is nearly a linear function of signal level. Bear in mind that this simple relation does not hold at the highest stimulus levels.

It is of interest to compare our fit to the Fletcher–Munson loudness data with other loudness-growth approximations in terms of their corresponding compression functions. This comparison is made in Fig. 3, which reproduces the compression estimates for the loudness data, along with the present fit, which is represented by the short dashed line. Steven's law suggests that loudness is proportional to stimulus intensity to the one-third power. This implies a constant compression of $\alpha = 3$ for all stimulus levels and is illustrated by the horizontal dotted line in Fig. 3. The Steven's law

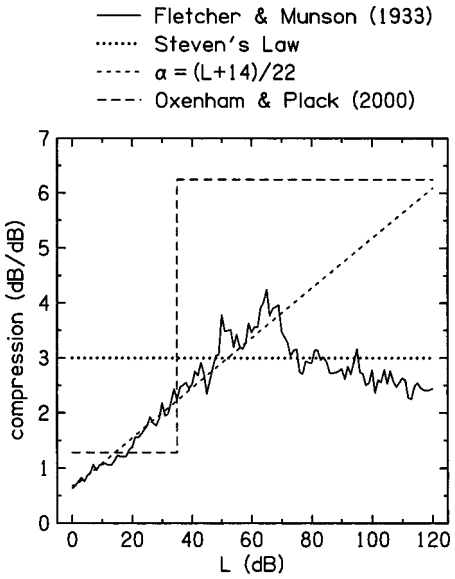


FIG. 3. Comparison of various compression estimates. The solid line was computed from the loudness data of Fletcher and Munson (1933). The dotted line represents a power-law nonlinearity that is the same at all levels, which is often called Steven's law. The short dashes represent the linear fit to loudness compression presented in this paper. The long dashes represent the two-slope peripheral nonlinearity used by Oxenham and Plack to fit psychophysical data.

approximation provides an estimate that agrees with the maximum compression estimate from the loudness data, but overestimates compression at threshold. Plack and Oxenham (1998) and more recently Oxenham and Plack (2000), suggest a two-slope approximation for the peripheral nonlinearity (at 6 kHz) that corresponds to $\alpha = 1.28$ for stimulus levels below 35 dB SPL and $\alpha = 6.25$ for levels above 35 dB SPL, as illustrated by the long dashed line in Fig. 3. This approach is consistent with the notion of there being less compression at threshold, compared to higher levels, but does not reflect the continuous relationship between amount of compression and level. Still, the nonlinear function suggested by Oxenham and Plack has been helpful in explaining psychophysical data.

The comparisons in Fig. 3 show the advantage below 70 dB of using a linear function to approximate the compression observed in the Fletcher–Munson loudness data. As stated earlier, the results imply that compression constantly increases from about 1 to about 4, so is not well described in this range as a constant, as in Steven's law, or by a step function, as suggested by Plack and Oxenham (1998). Functionally, it seems reasonable that the amount of compression should depend on stimulus level, increasing continuously with level. None of the simple functions shown in Fig. 3 is a satisfactory fit to the Fletcher–Munson loudness compression at all levels, nor do they completely describe the compression functions shown in Fig. 1. The simple compression functions are useful for their convenience; however, their limitations should be kept in mind.

C. Linear function and DPOAE growth

The successful fit to loudness compression by a linear function motivates the use of the same linear function to fit the DPOAE I/O data, as shown in Fig. 4. The upper panel

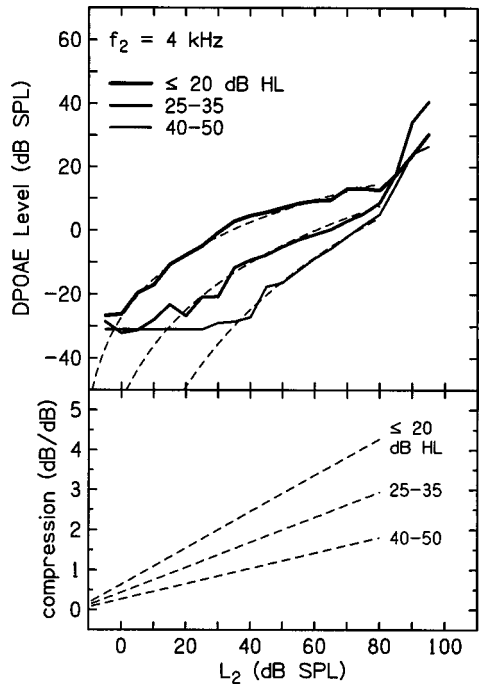


FIG. 4. DPOAE I/O functions for three hearing-loss groups. The solid lines are the same as shown in Fig. 1. The dashed lines represent fits to these data that have linear compression.

reproduces the median DPOAE I/O functions from Fig. 1 as solid lines. Again, data are shown for subjects with normal hearing and for subjects with different degrees of hearing loss. The bottom panel of Fig. 4 shows linear fits to the corresponding compression data that were shown in the bottom panel of Fig. 1. These fits were further constrained to intersect the x axis at the same L_2 value because this constraint allowed a simpler interpretation of the results while still providing a reasonable fit to the data. The linear fits were converted to the coordinate system shown in the upper panel of Fig. 4, where they are plotted as dashed lines. These lines, therefore, represent fits to the data that correspond to the linear compression functions shown in the lower panel. The x-axis intercept and slopes of the compression fits were determined by visual comparison of the corresponding I/O functions with the data.

The goodness of fit of the linear compression functions to actual DPOAE data is apparent in the upper panel of Fig. 4, at least for levels below about 80 dB SPL. The apparent disagreement between the data and the fits at low levels relates mainly to limitations inherent in DPOAE measurements. Recall that data collection ceased for any condition when the noise floor reached −25 dB SPL. With current technology, it is difficult to make reliable measurements of DPOAE levels lower than this value. This noise-floor limit applies regardless of whether hearing was normal or impaired. The departure between the fits and the actual data at low levels, therefore, may be related to the practical limits of DPOAE measurements. As stated earlier, there is some uncertainty in terms of what is reflected in high-level DPOAE measurements ($L_2 \geqslant 75$ dB SPL). Distortion measured at these high levels may reflect nonlinearities of a biological origin (perhaps cochlear or even middle ear) or may be contaminated by system distortion. As a result, the fits are limited to L_2 levels $\leqslant 80$ dB SPL.

The following compression function α_{d0} that we used to fit the DPOAE data from subjects with normal hearing (thresholds$\leqslant$20 dB HL) is identical to the one in Eq. (4) that we used to fit the Fletcher–Munson loudness data:

$$\alpha_{d0} = (L_2 + 14)/22. \tag{7}$$

The I/O relation that we used to fit the DPOAE level L_{d0} differs (by a constant) from the one that was used to fit the loudness data because L_{d0} is not zero when $L_2 = 0$

$$L_{d0} = 22 \cdot \ln((L_2 + 14)/48). \tag{8}$$

The mild (25–35-dB HL) and moderate (40–50-dB HL) hearing-loss groups were fit using compression functions (α_{d1} and α_{d2}) with the same intercept, but different slopes

$$\alpha_{d1} = (L_2 + 14)/32, \tag{9}$$

$$\alpha_{d1} = (L_2 + 14)/52. \tag{10}$$

Response level functions (L_{d1} and L_{d2}) that are consistent with these compression functions are shown in the upper panel of Fig. 4 and are defined by the following equations:

$$L_{d1} = 32 \cdot \ln((L_2 + 14)/74), \tag{11}$$

$$L_{d2} = 52 \cdot \ln((L_2 + 14)/87). \tag{12}$$

Note that the compression functions all intersect the x axis at $L_2 = -14$. The fact that it was possible to obtain a reasonable fit to all three groups (one with normal hearing and two with hearing loss) without changing the intercept of the compression function simplifies the characterization of hearing loss; however, the significance of this simplification is not clear. Only the compression slope of the fitted lines varies among the three hearing-loss groups. Human ears with normal hearing show the greatest amount of compression. The line fitted to the normal compression has the greatest slope. The amount of compression and the slope of the fitted line both decrease systematically as behavioral threshold increases.

In the upper panel of Fig. 4, it can be seen that the linear-compression curves fit the DPOAE I/O data almost as well as Eq. (1) fits the same data in Fig. 1, at least from the point where the DPOAE level rises above the noise floor to up to about 70–80 dB SPL. Equations (7)–(12) provide a simplified view of the effect of hearing loss on the DPOAE I/O function. Hearing loss elevates the DPOAE threshold and increases the slope of the DPOAE I/O function. These two effects can be modeled simply as a decrease in the slope of the DPOAE compression function. The similarity of the normal DPOAE I/O function to the normal Fletcher–Munson loudness function suggests that hearing loss may have an effect on loudness that is similar to the effect that is observed in DPOAE I/O functions. These results suggest a simple model for hearing loss as an increase in threshold that is accompanied by a change in the slope of the compression function.

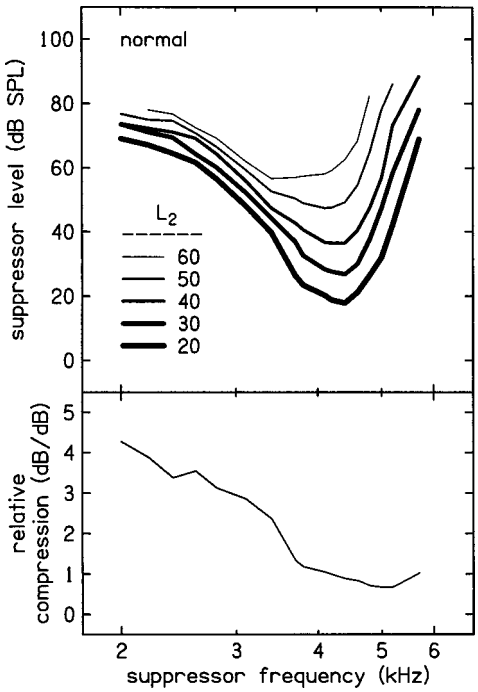


FIG. 5. DPOAE suppression tuning curves. The upper panel shows the level of a suppressor tone that was required to suppress DPOAE level by 3 dB as a function of frequency for six different primary tone levels. The lower panel shows an estimate of relative compression based on the vertical spacing of the curves in the upper panel.

D. Estimates of compression from DPOAE suppression tuning curves

Mean DPOAE suppression-tuning curves (STC) from Gorga *et al.* (2002) are shown in the upper panel of Fig. 5 for 13 subjects with normal hearing. Each STC represents the level of a third (suppressor) tone (f_3) that produced 3 dB of suppression, as a function of the suppressor frequency. The STCs in Fig. 5 were derived from measurements in which $f_2=4$ kHz, $L_2=20$, 30, 40, 50, and 60 dB SPL and $L_1=39+0.4\cdot L_2$.

The STCs have a larger vertical separation when the suppressor frequency is near 4 kHz ($f_3\approx f_2$) and a smaller separation when the suppressor frequency is near 2 kHz ($f_3<f_2$). At any specific frequency, the vertical separation is consistent across L_2 levels. Near 4 kHz, the STCs tend to be separated by about 10 dB, but they are only about 2.5 dB apart near 2 kHz. To quantify the vertical separation, we define the relative compression (RC) as the change in primary level L_2 divided by the change in suppressor level L_3 at suppression threshold (3 dB of suppression). The RC observed in the mean STCs is plotted in the lower panel of Fig. 5. Note that the RC is about 1 near 4 kHz, because the mechanical response to the primary tone and the suppressor tone at the 4-kHz place in the cochlea undergo the same compressive growth (see Fig. 1). The criterion amount of suppression (3 dB) is produced when the suppressor level approximates the primary level (L_2). The situation is different when the suppressor tone is near 2 kHz. The 2-kHz suppressor requires a higher level in order to produce suppression, but the mechanical response to the suppressor grows more linearly at the 4-kHz place, while the 4-kHz primary growth rate is compressive. Therefore, the relative growth rate reveals the compressive growth of the response to the primary tone. According to the mean RC curve (shown in the lower panel of Fig. 5), the response to the suppressor tone is growing about 4 times as fast as the response to the primary tone when the suppressor frequency is 2 kHz. If we assume that the 2-kHz suppressor response is growing linearly, then it follows that the primary response is undergoing a 4-to-1 compression.

Note that the maximum value of RC in Fig. 5 approximates the maximum compression of the DPOAE I/O function in the lower panel of Fig. 1 or Fig. 4. This similarity suggests that DPOAE suppression measurements with a low-frequency suppressor may offer an alternative method of estimating the compression associated with cochlear mechanical responses. The amount of compression can be extracted from an examination of how suppressor level to maintain a constant response (3 dB of suppression in the present example) grows with probe level (L_2) for an on-frequency and a low-frequency suppressor. This approach is similar to the behavioral approach taken by Oxenham and Plack (1997, 1998, 2000), and is an extension of the way Mills (1998) and Pienkowski and Kunov (2001) estimate a quantity related to cochlear-amplifier gain.

Gorga *et al.* (2003b) measured DPOAE STCs in ears with normal hearing and ears with hearing loss at 4 kHz. STCs could be measured for L_2 levels ranging from 20 to 70 dB SPL in ears with normal hearing, but were measurable for L_2 levels of only 50 to 70 dB SPL in ears with hearing loss. However, the STCs in normal and impaired ears were similar in shape at those absolute levels for which comparisons were possible. This suggests that compression estimates based on the STCs for impaired ears should be similar to those for normal-hearing ears. This observation seems to contradict our results based on DPOAE I/O functions, where compression estimates were noticeably less in impaired ears compared with normal ears. We are not yet able to explain this apparent contradiction.

E. DPOAE I/O and STC estimates of compression for individual subjects

All of the compression estimates shown above were based on group data. It is of interest to see whether estimates of cochlear compression based on DPOAE suppression tends to be more reliable than compression estimates based on the DPOAE I/O function. Figure 6 shows 38 individual compression estimates based on fitting logarithmic functions to DPOAE I/O functions from normal-hearing ears, as illustrated in Fig. 4 and Eqs. (7) and (8). As in Fig. 3, the compression functions were all required to intersect the *x* axis at -14 dB. The estimated compression at 65-dB SPL varies from 1.75 to 7.58. The relative-compression functions in Fig. 7 are based on DPOAE suppression measurements in 12 normal-hearing ears. At 2.2 kHz, the compression estimates vary from -0.35 to 5.72.

We would not expect the actual cochlear compression in normal-hearing ears to vary as much as indicated in Figs. 6 and 7. Thus, it appears that neither method is sufficiently robust to provide reliable estimates of cochlear compression

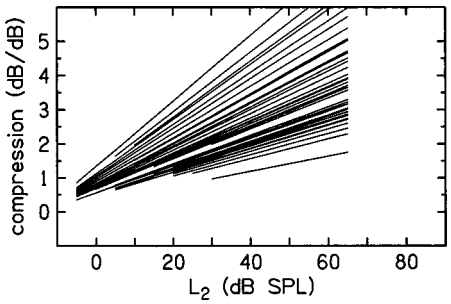


FIG. 6. Compression estimates for individual normal-hearing ears based DPOAE I/O functions. These results are comparable to the lower panel of Fig. 4.

in individual ears. This is a disappointing observation because it means that the present approach cannot be used to predict compression in individual subjects. Some of the observed variability may be due to our DPOAE measurement methods. For example, *in-the-ear* stimulus pressure calibration is known to depend on the insertion depth of the probe–microphone (Neely and Gorga, 1998), and our choice of stimulus level combinations will not always produce the largest possible DPOAE levels in every subject. It is also possible that DPOAE levels do not always provide an accurate indication of the growth of cochlear responses that are directly relevant to hearing. One goal of future work is to gain a better understanding of the cause(s) of the variability in our individual compression estimates.

IV. CONCLUSIONS

When $L_1 = 39 + 0.4 \cdot L_2$, the DPOAE level versus L_2 curve for normal-hearing ears is similar in shape to the Fletcher–Munson (1933) loudness curve. This suggests that the same peripheral nonlinearity determines the compressive growth of both of these two I/O functions.

When the Fletcher–Munson log-loudness data or the DPOAE I/O data are fit by a logarithmic response function, then the compression, defined as the reciprocal of the slope of the response function, is a nearly linear function of stimulus level. This observation leads to a simple relation between loudness and stimulus level at low and moderate levels.

DPOAE I/O data and DPOAE suppression data from groups of normal-hearing ears provide similar estimates of the maximum cochlear compression; however, individual I/O and suppression measurements apparently do not provide robust estimates of compression.

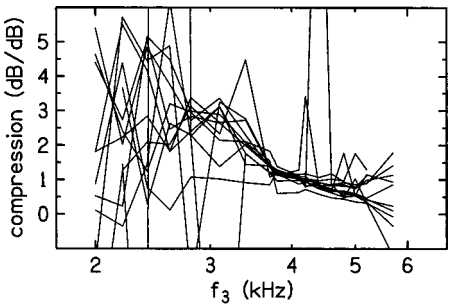


FIG. 7. Relative compression estimates for individual normal-hearing ears based on suppression tuning curves. These results are comparable to the lower panel of Fig. 5.

ACKNOWLEDGMENTS

This work was supported by a grant from the NIH (NIDCD R01 2251). Portions of this work were presented at the Twenty-Fifth Annual Midwinter Research Meeting of the Association for Research in Otolaryngology.

Allen, J. B., and Neely, S. T. (**1992**). "Micromechanical models of the cochlea," Phys. Today **45**, 40–47.

ANSI (**1996**). ANSI S3.6-1996, "Specifications for Audiometers" (American National Standards Institute, New York).

Arnold, D. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1999**). "High-frequency hearing influences lower-frequency distortion-product otoacoustic emissions," Arch. Otolaryngol. Head Neck Surg. **125**, 215–222.

Boege, P., and Janssen, T. (**2002**). "Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O functions in normal and cochlear hearing loss ears," J. Acoust. Soc. Am. **111**(4), 1810–1818.

Brownell, W. E. (**1990**). "Outer hair cell electromotility and otoacoustic emissions," Ear Hear. **11**(2), 82–92.

Buus, S., Musch, H., and Florentine, M. (**1998**). "On loudness at threshold," J. Acoust. Soc. Am. **104**, 399–410.

Dallos, P. J., and Harris, D. M. (**1978**). "Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level," Hear. Res. **49**, 225–246.

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., and Gorga, M. P. (**2001**). "Distortion product otoacoustic emission input/output functions in normal and impaired ears," J. Acoust. Soc. Am. **110**, 3119–3131.

Dorn, P. A., Piskorski, P., Gorga, M. P., Neely, S. T., and Keefe, D. H. (**1999**). "Predicting audiometric status from distortion product otoacoustic emissions using multivariate analyses," Ear Hear. **20**, 345–362.

Fletcher, H., and Munson, W. (**1933**). "Loudness: its definition, measurement, and calculation," J. Acoust. Soc. Am. **9**, 1–10.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Konrad-Martin, D. (**2002**). "The use of DPOAE suppression as an estimate of response growth," J. Acoust. Soc. Am. **111**, 271–284.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Hoover, B. M. (**2003a**). "Further efforts to predict pure-tone threshold from distortion-product otoacoustic emission input/output functions," J. Acoust. Soc. Am. **113**, 3275–3286.

Gorga, M. P., Neely, S. T., Dierking, D., Dorn, P. A., Hoover, B. M., and Fitzpatrick, D. (**2003b**). "Distortion-product otoacoustic emission tuning curves in normal and hearing-impaired ears," J. Acoust. Soc. Am. (in press).

Harris, F. P., and Probst, R. (**2002**). "Otoacoustic emissions and audiometric outcomes," in *Otoacoustic Emissions: Clinical Applications*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme Medical, New York), pp. 213–242.

Kiang, N. Y.-S., Liberman, M. C., and Levine, R. A. (**1976**). "Auditory nerve activity in cats exposed to ototoxic drugs and high-intensity sounds," Ann. Otol. Rhinol. Laryngol. **85**, 752–768.

Kimberley, B. P., Hernandi, I., Lee, A. M., and Brown, D. K. (**1994**). "Predicting pure tone thresholds in normal and hearing-impaired ears with distortion product emissions and age, Ear Hear. **15**, 199–209.

Kimberley, B. P., Brown, D. K., and Allen, J. B. (**1997**). "Distortion product emissions and sensorineural hearing loss," in *Otoacoustic Emissions: Clinical Applications*, edited by M. S. Robinette and T. J. Glattke (Thieme Medical, New York), pp. 181–204.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). "Level and growth behavior of the 2 $f1-f2$ distortion product otoacoustic emission and its relation to auditory sensitivity in normal hearing and cochlear hearing loss," J. Acoust. Soc. Am. **103**, 3431–3444.

Liberman, M. C., and Dodds, L. W. (**1984**). "Single-unit labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," Hear. Res. **16**, 55–74.

Martin, G. K., Probst, R., and Lonsbury-Martin, B. L. (**1990**). "Otoacoustic emissions in human ears: Normative findings," Ear Hear. **11**, 106–120.

Mills, D. M. (**1998**). ''Interpretation of distortion product otoacoustic emission measurements. II. Estimating tuning characteristics using three stimulus tones,'' J. Acoust. Soc. Am. **103**(1), 507–523.
Neely, S. T., and Gorga, M. P. (**1998**). ''Comparison between intensity and pressure as measures of sound level in the ear canal,'' J. Acoust. Soc. Am. **104**, 2925–2934.
Neely, S. T., and Kim, D. O. (**1986**). ''A model for active elements in cochlear biomechanics,'' J. Acoust. Soc. Am. **79**, 1472–1480.
Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). ''A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **110**, 2045–2064.
Norton, S. J., Gorga, M. P., Widen, J. E., Folsom, R. C., Sininger, Y., Cone-Wesson, B., Vohr, B. R., Masher, K., and Fletcher, K. A. (**2000a**). ''Identification of neonatal hearing impairment: Evaluation of transient evoked otoacoustic emission, distortion product otoacoustic emission, and auditory brain stem response test performance,'' Ear Hear. **21**(5), 508–528.
Norton, S. J., Gorga, M. P., Widen, J. E., Folsom, R. C., Sininger, Y., Cone-Wesson, B., Vohr, B. R., and Fletcher, K. A. (**2000b**). ''Identification of neonatal hearing impairment: Summary and recommendations,'' Ear Hear. **21**(5), 529–535.
Oswald, J. A., Müller, J., and Janssen, T. (**2002**). ''Audiometric threshold estimation in cochlear hearing loss ears by means of weighted extrapolated DPOAE I/O functions,'' Abstract 774, Twenty-fifth Midwinter Meeting of the Association for Research in Otolaryngology.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Oxenham, A. J., and Plack, C. J. (**1998**). ''Suppression and the upward spread of masking,'' J. Acoust. Soc. Am. **58**, 867–869.
Oxenham, A. J., and Plack, C. J. (**2000**). ''Effects of masker frequency and duration in forward masking: Further evidence for the influence of peripheral nonlinearity,'' Hear. Res. **150**, 258–266.
Pienkowski, M., and Kunov, H. (**2001**). ''Suppression of distortion product otoacoustic emissions and hearing threshold,'' J. Acoust. Soc. Am. **109**, 1496–1502.
Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Robles, L., and Ruggero, M. A. (**2001**). ''Mechanics of the mammalian cochlea,'' Physiol. Rev. **81**, 1305–1352.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Ruggero, M. A., Rich, N. C., Narayan, S. S., and Robles, L. (**1997**). ''Basilar membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Stelmachowicz, P. G., Dalzell, S., Peterson, D., Kopun, J., Lewis, D. L., and Hoover, B. E. (**1998**). ''A comparison of threshold-based fitting strategies for nonlinear hearing aids,'' Ear Hear. **19**, 131–138.
Stover, L., Gorga, M. P., Neely, S. T., and Montoya, D. (**1996**). ''Toward optimizing the clinical utility of distortion product otoacoustic emission measurements,'' J. Acoust. Soc. Am. **100**, 956–967.

Effects of interaural frequency difference on binaural fusion evidenced by electrophysiological versus psychoacoustical measures[a)]

Jianxun Zhou[b)] and John D. Durrant[c)]
Departments of Communication Science and Disorders and Otolaryngology, University of Pittsburgh, Forbes Tower 4033, Pittsburgh, Pennsylvania 15260

(Received 9 October 2002; revised 23 May 2003; accepted 17 June 2003)

The binaural interaction component (BIC=sum of monaural−true binaural) of the auditory brainstem response appears to reflect central binaural fusion/lateralization processes. Auditory middle-latency responses (AMLRs) are more robust and may reflect more completely such binaural processing. The AMLR also demonstrates such binaural interaction. The fusion of dichotically presented tones with an interaural frequency difference (IFD) offers another test of the extent to which electrophysiological and psychoacoustical measures agree. The effect of IFDs on both the BIC of the AMLR and a psychoacoustical measure of binaural fusion thus were examined. The perception of 20-ms tone bursts at/near 500 Hz with increasing IFDs showed, first, a deviated sound image from the center of the head, followed by clearly separate pitch percepts in each ear. Thresholds of detection of sound deviation and separation (i.e., nonfusion) were found to be 57 and 209 Hz, respectively. However, magnitudes of BICs of the AMLR were found to remain nearly constant for IFDs up to the 400-Hz (limit of range tested), suggesting that the AMLR-BIC does not provide an objective index of this aspect of binaural processing, at least not under the conditions examined. The nature of lateralization due to IFDs and the concept of critical bands for binaural fusion are also discussed. Further research appears warranted to investigate the significance of the lack of effect of IFDs on the AMLR-BIC. Finally, the IFD paradigm itself would seem useful in that it permits determination of the limit for nonfusion of sounds presented binaurally, a limit not accessible via more conventional paradigms involving interaural time, phase, or intensity differences. [DOI: 10.1121/1.1600718]

PACS numbers: 43.64.Qh, 43.64.Ri, 43.66.Pn [WPS]

I. INTRODUCTION

The binaural interaction component (BIC) is the difference waveform between the true binaural evoked response and the sum of the monaurally evoked responses (Dobie and Norton, 1980; Levine, 1981). It has been observed that binaurally evoked responses are always somewhat smaller than the sum of the two monaurally evoked responses. A difference potential (i.e., the BIC) is obtained by subtracting the true binaural potential from the sum of the monaurally evoked potentials (Dobie and Norton, 1980; Levine, 1981; Wrege and Starr, 1981; Furst *et al.*, 1985; Ito *et al.*, 1988; McPherson and Starr, 1993). Several possible artifacts that might produce such a difference potential have been ruled out, such as middle-ear reflex effects (Levine, 1981; Ito *et al.*, 1988), acoustic cross talk (Ainslie and Boston, 1980; Levine, 1981; Furst *et al.*, 1985), and right-ear versus left-ear asymmetry (Decker and Howe, 1981; Spivak and Seitz, 1988; McPherson and Starr, 1993). The validity of BIC measurement as a potential objective evaluation of the binaural system has been suggested by results of analyses of three-channel Lissajous' trajectories (or 3-CLTs) (Polyakov and Pratt, 1995). Evoked potential changes in the three-dimensional voltage space for auditory brainstem responses (ABRs) and auditory middle-latency responses (AMLRs) demonstrated trajectories specific to the BIC, suggesting unique dipole generators.

It also has been demonstrated that the BIC of ABR correlates with lateralization and fusion. Furst *et al.* (1985) investigated the relationship between the changes in both perception of lateralization and BIC amplitudes of responses evoked by dichotic clicks with interaural time and intensity differences (ITDs and IIDs, respectively). Their results showed that whenever the BIC was clearly present, the clicks were perceived as a unitary fused image; whenever the clicks were perceived as not fused, the BIC was undetectable. The authors suggested that the BIC is a physiological correlate of binaural fusion and is generated by a brainstem structure essential for the perception of click lateralization. McPherson and Starr (1995) also found that the presence of a BIC of the ABR to be correlated with the fusion of binaural click stimuli, and the amplitudes of the BIC correlated negatively with the degrees of lateralization of the intracranial images.

Another approach to the study of the binaural system is the fusion and lateralization of sounds with interaural frequency differences (IFDs). When the IFD is slightly different from zero, a tone is perceived that seems to rotate in the head at a rate equal to the IFD (Durlach and Colburn, 1978). This

[a)]Portions of this report were presented at the XVII Biennial Symposium of the International Evoked Response Audiometry Study Group, 22–27 July 2001, Vancouver, BC, Canada.

[b)]Current address/affiliation: Kresge Hearing Research Institute, University of Michigan, Ann Arbor, MI 48109-0506.
[c)]Electronic mail: durrant@csd.pitt.edu

is the binaural beat, which is mainly observed with relatively low-frequency pure tones. As the IFD is made larger and larger, the sound image tends to become diffuse and finally splits into two smooth tones of separate pitches, that is, one in each ear.

To further evaluate the extent to which the BIC reflects binaural processing, it seemed useful to consider whether the perceptual aspects of binaural lateralization and fusion caused by IFD can be described by BICs. Fowler (1989) investigated the effect of bifrequency binaural stimulation on the BICs of the ABR. When 1-ms tone bursts presented to the two ears with a 2000-Hz IFD (one ear with 1000 Hz and the other with 3000 Hz), the BIC was still recorded. However, there were no psychoacoustical data provided to characterize the subject's perception (fusion or not). In any event, the effective short duration of stimulation of the ABR makes it difficult to critically evaluate behavioral responses due to broad spectral splatter, thus making it difficult to judge the stimulus in terms of pitch. The AMLR is inherently more robust and of longer latency, and can be expected to reflect binaural processing more extensively (i.e., from anatomical considerations). Stimuli with longer durations, in turn, can be employed readily in eliciting AMLRs (Hall, 1992). Again, BICs also have been observed in the AMLR, yielding larger amplitudes compared to BICs of the ABR (Dobie and Norton, 1980; McPherson and Starr, 1993).

It thus seemed likely that a common electrophysiological and psychoacoustical paradigm could be designed to permit direct comparisons between BICs of the AMLR and perceptions of IFDs. On the one hand, relatively brief stimuli of short rise times are important for good neural synchrony, to optimize AMLR amplitude, and minimal overlap with the response epoch, to minimize stimulus artifact. On the other hand, good tonal quality requires sufficiently long durations. Optimization of the two stimulus paradigms, consequently, is somewhat at odds. Mark and Rattay (1990) found that the threshold for the perception of click pitch of a stimulus centered on 500 Hz was about 19 ms. In addition, it was shown by results of their study that, although very short acoustic signals (1-to-3-cycle sinusoidal pulses) were recognized as clicks, they still could be distinguished by pitch. Stimuli of durations from a click to 50 ms elicit AMLRs of robust magnitude (Hall, 1992), but durations of more than 19 ms or so would potentially cause significant interference from artifact. A reasonable compromise, from all considerations, might be a stimulus duration around 20 ms.

In addition to possible research interests, it is of potential clinical interest to develop objective methods to evaluate central auditory processing. The capacity to fuse sounds with IFDs is implied potentially by the clinical binaural fusion test, wherein speech is filtered such that the high-pass (high-frequency) portion is presented to one ear and the low-pass portion is presented to the other ear. Although neither ear alone has enough information for precise word identification, normal-hearing subjects can readily recognize a fused speech signal. However, patients with central nervous system dysfunction may have poor word recognition and/or cannot fuse the two signals. Whether such subjects have difficulty fusing sounds with IFDs remains to be seen, but seems likely. Still, further work is needed before embarking upon clinical studies, namely the evaluation of AMLR-BICs in neurologically intact subjects.

Consequently, this study was dedicated to testing the hypothesis that the magnitude of the AMLR-BIC reflects the limit of interaural frequency difference for binaural fusion. Support of this hypothesis would further support the idea of an intimate relationship between this component and binaural processing underlying binaural fusion or lateralization. Practically, the use of the BIC AMLR in an IFD paradigm might make the BIC more clinically attractive if, as anticipated, a robust BIC were obtained. The availability of directly comparable electrophysiological and psychoacoustical responses, measured in directly corresponding paradigms, is also attractive as a potential research tool. Since low-frequency signals have been thought to contribute mainly to the binaural beats of IFDs and have been extensively studied in the literature, effects of IFDs around 500 Hz were the focus of this study.

II. METHOD

A. Subjects

Ten normal-hearing subjects (six females and four males, ranging in age from 21 to 30 years) volunteered to participate in this study. Subjects had negative otologic histories, pure-tone hearing thresholds within normal limits (mean thresholds less than 10 dB hearing level [HL]), and symmetric hearing that was verified psychophysically, as follows: when identical stimuli were presented binaurally, all subjects perceived a fused sound image at midline. Informed consent was obtained from each subject in this protocol approved by the Internal Review Board of the University of Pittsburgh for the use of human subjects.

B. Stimuli

All the tonal stimuli were generated and controlled by a commercial hardware and software package (Tucker-Davis Technologies and Siggen, respectively) via either a one- or two-channel paradigm. Output signals were sent to insert earphones (ER-3A) via the TDT system's earphone amplifier module. Tone bursts of 500 Hz with a total duration of 20 ms were gated using a Blackman window and presented at 65-dB sensation level (SL). This type of tone bursts served psychoacoustical and electrophysiological tests equally. In addition, tone bursts with durations of 250, 100, and 60 ms also were generated for training purposes (*vis-à-vis* the psychoacoustical part of the study). Tone bursts of 250 ms also were employed in the psychoacoustical component of the study to provide a basis of comparison for results obtained using brief stimuli appropriate to AMLR analysis, i.e., in the main part of the experiment. The longer tone bursts provide more tonal sensations and are more characteristic of stimuli employed in past psychoacoustic studies of IFDs.

C. Psychoacoustical protocol

Thresholds were determined for binaural nonfusion (a clear perception of separate tones in the two ears) and just-

A.

PARADIGM	Binaural non-fusion		JDD	
CONDITION (units of measure)	250-ms (Hz)	20-ms (Hz)	IFD (Hz)	IPD (Degrees)
THRESHOLD	36 +/-17	209 +/- 98	59 +/- 39	47 +/- 23

B.

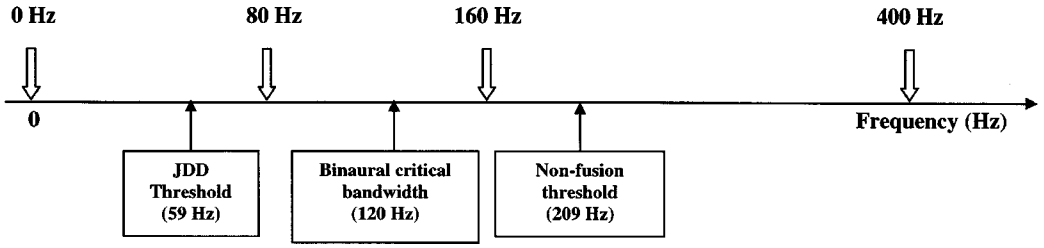


FIG. 1. Psychoacoustical results and interrelations with electrophysiological paradigms. (A) Psychophysical thresholds for different test conditions: Binaural nonfusion for 250- and 20-ms tone bursts; just detectable deviation (JDD) for 20-ms tone bursts with interaural frequency difference (IFD or Δf) and interaural phase difference (IPD). (B) Illustration of the rationale for choosing IFD values in the electrophysiological study. Horizontal right-facing line represents the frequency difference between the two ears (Δf). Upward arrows represent the values of Δf to which different thresholds (indicated in the boxes below) correspond. The values of binaural nonfusion and JDD were preselected based upon pilot studies, which were close to their actual values (indicated in the parentheses, 59 and 209 Hz, respectively). The threshold of binaural critical bandwidth (120 Hz) was estimated from literature (Scharf *et al.*, 1976; Yama and Small, 1983). The actual Δf's employed in this study (down arrows) therefore may theoretically sample each category.

detectable deviation (JDD) of sound image from the center of the head (tested for 20-ms tone bursts only). Subjects were seated in a sound isolation booth and instructed to register their response on an answer sheet. Binaural nonfusion thresholds, again, were tested using both 250-ms and 20-ms stimuli with interaural frequency differences (IFDs or Δf's); JDDs were tested using 20-ms stimuli with both IFDs and interaural phase differences (IPDs, or $\Delta\varphi$'s). The criterion for binaural nonfusion, to iterate, was the clear perception of two sound images of different pitches, one in each ear. However, subjects were not required to identify which ear received the higher versus lower pitch stimulus. The method of constant stimuli was used for determining thresholds for evaluating both binaural nonfusion (IFD paradigm) and sound-image deviation (IPD paradigm). The reason for using this psychophysical method, rather than a forced-choice method, was that multiple sound perceptions might exist for short-duration signals with increasing IFD. A deviated sound-image and/or ''chord'' (or distorted) perception (Thurlow and Bernstein, 1957) might occur before the signal is perceived to split into two smooth images (tones) with separate pitches in each ear. Subjects probably could use these cues, rather than acutal nonfusion, to make a judgment. This problem is expected to be especially prominent in the forced-choice procedure since the focus of this method is ostensibly on detection of differences, *per se*.

During the threshold tests, a warning signal (a burst of white noise) was presented 500 ms before each pair of target tone bursts. Subjects were told to pay attention only to the second test tone, and they had 2.5 s to respond. For binaural nonfusion tests, subjects were asked to circle ''yes'' on the answer sheet only if they could clearly detect two distinct tone bursts. Otherwise, they circled ''no.'' For JDD tests, subjects were asked to circle ''yes'' simply if they found a sound image deviated from the center of the head.

Actual pairs of target tone bursts during each trial were chosen from a pilot study, ensuring that the thresholds of interest would fall roughly in the middle of test ranges. Briefly, during the nonfusion task for 250- and 20-ms tone bursts, ΔFs were chosen as 0, 15, 30, 60, 120, 240 Hz; and 0, 80, 160, 240, 320, 400, and 480 Hz, respectively; during the JDD task for IFD and IPD, ΔFs ($\Delta\varphi$s) were chosen as 0, 10, 20, 40, 80, 160 Hz, and 0, 7.5, 15, 30, 60, 90 deg, respectively. Ten judgments were repeated for each trial, randomized by sequence and ear. To ensure minimal response bias, the stimulus test pairs were manipulated, based on pilot data, such that nearly half of them most likely would fall below threshold and the other half above threshold, i.e., the subject should have, overall, nearly equal opportunity to perceive a difference (or not) across trials on a chance basis alone. The threshold value was determined using MATLAB (MathWorks, Inc.) to fit the data to a normal cumulative distribution curve via the method of least squares. The values corresponding to the 50th percentile of the ''yes'' responses in the cumulative distribution were defined as thresholds. It thus was assumed that the response rate follows the normal cumulative distribution function; practically, fitting this function to the data helped also to minimize the influence of any

outlier responses. Before these ''formal'' tests, subjects were trained for the evaluation of perception of binaural nonfusion by presenting them sequentially signals from $\Delta F = 0$ to the largest value tested. Signals of 250-ms duration were presented first, then 100, 60, and 20 ms, in effect in order of most to least pitch strength. In addition, subjects were asked to describe their perceptions of the sounds presented to assure the examiner of the subject's complete comprehension of the task.

D. Electrophysiological protocol

Identical 20-ms tone bursts to those used in the psychoacoustical study were used to obtain auditory evoked potentials. A total of four pairs of ΔFs (0, 80, 160, and 400 Hz) between the two ears was investigated, so the possible pairs (right/left or left/right, in a random order) were: 500/500, 540/460, 580/420, 700/300, respectively. These frequency separations were chosen from the pilot study according to the JDDs and thresholds of binaural nonfusion observed, as well as published binaural critical bandwidths (Scharf *et al.*, 1976; Yama and Small, 1983). The rationale for the choice of ΔFs in the evoked potential study is illustrated in Fig. 1(B).

Auditory evoked potentials were recorded between electrodes at Cz (noninverting) and Cvii (the 7th cervical vertebra) (Polyakov and Pratt, 1995). A ground electrode was placed on the chin. Electrode resistances were balanced and below 3 KΩ. Tone bursts of fixed phase (condensation [chosen arbitrarily]) were presented at 11.1/s and delivered through earphones (ER-3A, same as those used in the psychoacoustic study). Auditory evoked potentials were recorded for each of three conditions: left monaural, right monaural, and binaural. The analysis window was 85 ms. Responses were averaged over 512 sweeps; this was repeated three times. The ear of presentation and frequency separation were randomly selected. Analog prefiltering was used with a bandwidth of 10 to 1500 Hz (12 dB/oct). The averaged evoked responses were stored digitally (256 data points per sweep) for further analysis off-line.

A digital filter (Butterworth) generated in MATLAB (MathWorks, Inc.) with a bandpass of 8–120 Hz was applied to all recorded responses during off-line analysis. This provided further reduction of high-frequency noise and stimulus artifact, as well as low-frequency noise or drift. To further increase the signal-to-noise ratio, the two waveforms most alike (judged by visual examination) of the three repeated responses for each condition were selected for final averaging together.

The BIC was determined by subtracting the binaurally evoked potentials from the sum of the monaurally evoked potentials for each IFD (Fig. 2). Peak-to-peak amplitudes of Na–Pa and Pa–Nb were calculated for monaural and binaural AMLRs. The BICs of the AMLR demonstrated very similar wave components as those of the monaural responses, so amplitudes were measured the same way, but for clarity the peaks are labeled as Na′, Pa′, and Nb′, respectively (Fig. 2). Latencies were corrected by subtracting 0.9 ms to compensate for the ER-3A tubal insert earphones used in the study (Hall, 1992). To illustrate central tendencies, grand average waveforms were obtained by averaging across all subjects

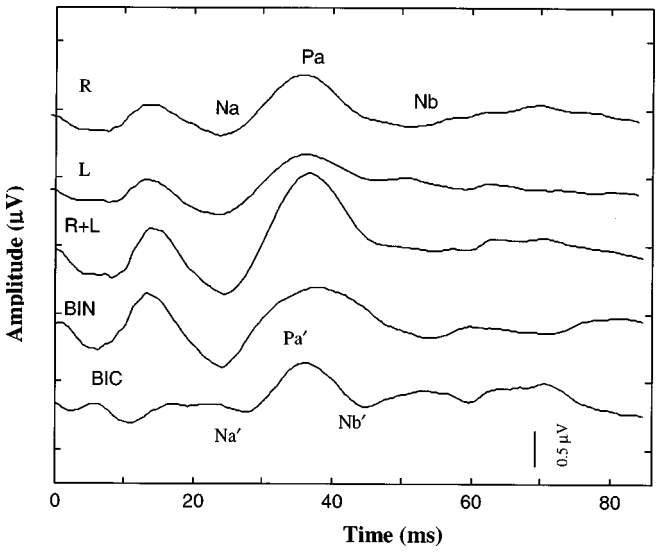


FIG. 2. Examples of filtered auditory middle latency responses (AMLRs) and binaural interaction component (BIC). Figure shows the grand average of monaural responses (R and L), sum of the monaural responses (R+L), the binaural responses (BIN), and the derived BIC.

for each condition. Otherwise, means and standard deviations from discrete peak latencies were used for graphical and statistical analysis, which in turn was a multivariate analysis of variance (MANOVA), in deference to the multiple treatments and repeated measures used with the same subjects.

III. RESULTS

A. Psychoacoustical study

During training sessions, when the IFD was increased, all subjects reported that percepts of the 250-ms tone bursts changed progressively (as expected) from a fused center image to binaural beats, followed by the perception of ''roughness,'' and finally two distinct pitches, one in each ear. For 20-ms tone bursts, when the IFD increased from zero to the largest value tested (480 Hz), all subjects first described a single compact sound image in the center of the head, followed by the perception of a single compact tone moving to one ear or in the back of the head, then a diffuse sound image characterized by such verbal descriptors as ''noncompact,'' ''distorted,'' ''unbalanced,'' or ''strange.'' Finally, all subjects clearly detected two sounds of distinct and separate pitches in each ear for the largest IFDs. Still, only two subjects could indicate readily which ear had the higher versus lower pitch (although this was not a requirement for the nonfusion threshold measurement, as noted earlier). However, some subjects' responses to 250-ms tone bursts displayed apparent deviations from the typical pattern of behavior. Examination of these data indicated that judgments were inconsistent with their reports during training, and might have been biased by binaural beats during the experiment, despite training. Their data were excluded from the final analysis.

Ultimately, most subjects showed reasonably normally distributed response rates, justifying the curve fitting applied, and the mean threshold of binaural nonfusion for the 250-ms tone bursts was found to be 36 ± 17 Hz [Fig. 1(A)]. For

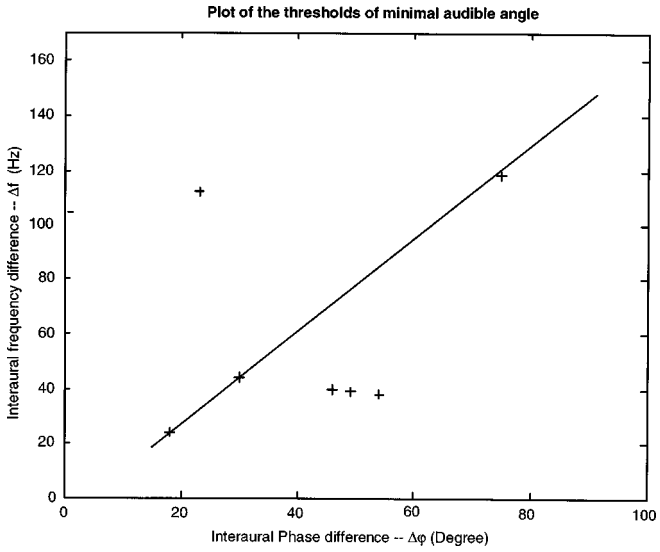


FIG. 3. Plot of the thresholds of just detectable deviation (JDD) from interaural frequency difference (Δf) as a function of those from interaural phase difference ($\Delta\phi$).

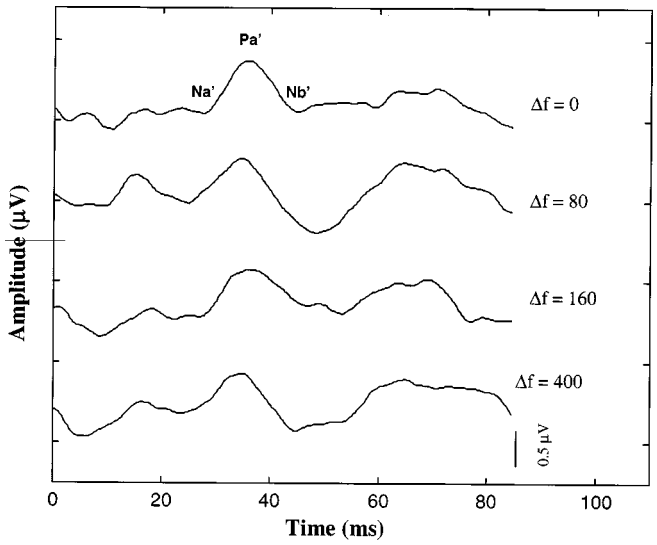


FIG. 4. Grand averages of binaural interaction components (BICs) for different interaural frequency differences (Δf):$\Delta f = 0$, 80, 160, and 400 Hz. Robust BICs observed for all Δf's without systematic effects observed upon inspection.

20-ms tone bursts, all subjects demonstrated responses consistent with results observed during training and appeared (upon inspection) to be normally distributed responses. The threshold for nonfusion, again, determined via curve fitting, dramatically increased, as expected for the increased spectral splatter of this stimulus, namely 209 ± 98 Hz. Still, this threshold fell well below the highest IFD tested.

Theoretically, IFDs produce concomitant interaural phase differences, potentially influencing the percept of the fused image (for example, its laterality and/or compactness). Evaluated with brief stimuli, the JDD as a function of IFD was found to be 59 ± 39 Hz. The JDD was also determined for interaural phase differences with no IFD and found to be 47 ± 23 deg for these brief stimuli. The results revealed a potential covariance of the two measures, albeit imperfect on a subject-by-subject basis (Fig. 3). (On the other hand, this was not the focus of this study.)

B. Electrophysiological study

Most recorded monaural and binaural responses showed a consistent and well-identified Na–Pa–Nb complex, especially for the Na–Pa component. Sometimes an additional positive peak, Pb, and a negative peak, Nc, was observed but varied across subjects and among stimulus conditions. Therefore, only measurements based upon the three major components, Na, Pa, and Nb, were submitted to further analysis. The BIC was observed in all four IFD conditions and consistently demonstrated very similar wave components as those of the monaural/binaural responses (again, for clarity, labeled Na′, Pa′, and Nb′, respectively; Figs. 2 and 4). The Na′–Pa′ magnitude, indeed, showed a generally flat pattern as a function of ΔF, as illustrated in Fig. 5(A). This was largely true for Pa′–Nb′ as well, except for a slight increase in amplitude at $\Delta F = 80$ Hz [Fig. 5(A)] However, neither measure showed significant differences among the four ΔFs tested ($\alpha = 0.05$). In Fig. 5(B), magnitudes are calculated as a percentage *re*: the summed monaural response, but the pattern remains essentially the same, reflecting no systematic change (absolute or relative) over IFD that might be related to the psychacoustic results. Figure 5(C) presents peak-to-peak measures of binaural responses as a function of ΔFs. None of these components displayed a sudden amplitude increase, that is, as if a critical ΔF had been reached. The latencies, peak-to-peak measures, and percentage ratios of BIC to the sum of the monaural responses for each test condition can be found in Table I.

IV. DISCUSSION

The results of this study do not support the hypothesis that the BIC of AMLR reflects binaural processing underlying binaural fusion versus lateralization of sounds of different frequency between ears. As noted in the Introduction, such fusion is known to occur over a critical bandwidth. Stimulation with an IFD within this band can be viewed as effectively one frequency component having a dynamic interaural phase difference (Robinson, 1971). The mean phase difference can be shown to be directly proportional to ΔF. As seen in our results on IFD versus IPD (Fig. 3), the concomitant phase difference is probably the main factor determining fusion versus nonfusion and, presumably, the nature of the resulting auditory image. The imperfect relationship between IFD and IPD in this study is most likely due to the limited sample of the subjects' performance, the relatively small subject sample size, and/or possibly less precise measures of lateralization employed in this study. (A true centered image presented as a reference in each trial might help subjects make a more accurate judgment.) Since the amplitudes of the BIC have been demonstrated to correlate inversely with the degree of lateralization of the intracranial images (McPherson and Starr, 1995), it might be interesting to ask if there are also inverse linear relationships between the BIC and lateralizations caused by either the IPD or IFD paradigm within a critical band.

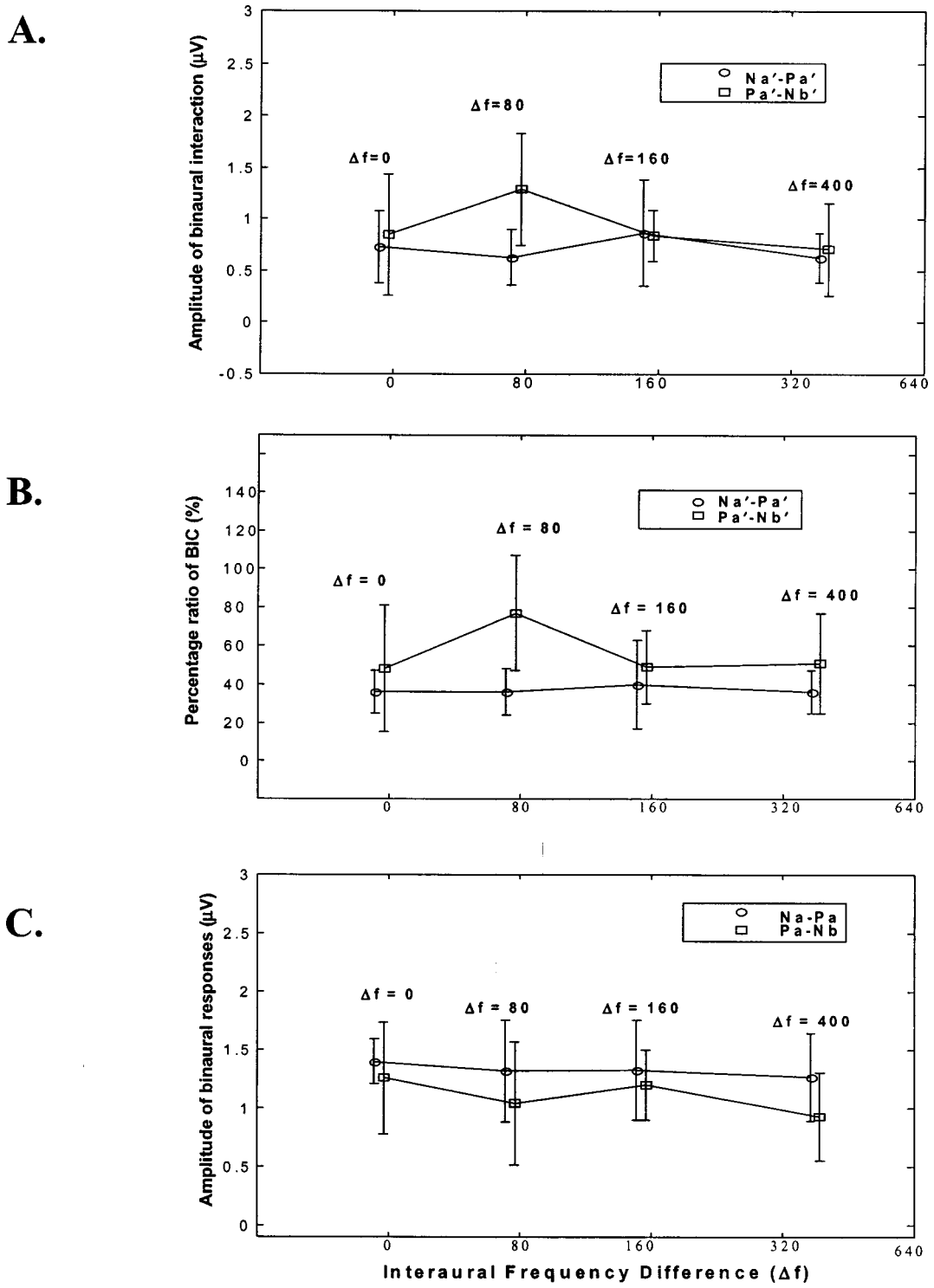


FIG. 5. (A) Amplitudes of binaural interaction as a function of interaural frequency difference (Δf) (Circle: Na′–Pa′; Square: Pa′–Nb′). The Na′–Pa′ magnitude, indeed, showed a generally flat pattern as a function of Δf. This was largely true for Pa′–Nb′ as well, except for a slight increase in amplitude at $\Delta f = 80$ Hz. Neither measure showed statistical differences among IFDs ($P > 0.05$). (B) Percentage ratios of binaural interaction to the sum of the monaural responses as a function of Δf. Again, no obvious changes over Δf's (Circle: Na′–Pa′ ratio; Square: Pa′–Nb′ ratio). (C) Amplitude of binaural responses as a function of Δf (Circle: Na–Pa; Square: Pa–Nb). This flat pattern is consistent with the psychophysical data (Scharf, 1969; see the text for details). [Data graphed: means +1 standard deviation.]

Since the BIC is a difference potential derived from the auditory evoked potentials, it can be viewed alternatively as a synchronized or onset response. If this is true, the BIC, as an index of binaural onset response, might not be sensitive to the binaural processing of IFD, in which longer processing time may be needed, especially if the behavioral response depends substantially upon temporal encoding. This issue also may underlie the discrepancy between results of this study and others using the lateralization paradigm based on IID and ITD. Nevertheless, the IFD paradigm provides potentially a new useful experimental probe of central auditory function. A substantial advantage of the IFD manipulation in

TABLE I. Mean binaural interaction component latencies, peak-to-peak amplitudes, and percentage ratios of binaural interaction to the sum of monaural responses. The right most column listed the percentage ratios of three major binaural interaction components to the sum of the monaural responses obtained from McPherson and Starr (1995). All values expressed as mean (±1 standard deviation).

	Latency (ms)			Ampltude (vU)		Percentage ratio of BIC		
Δf (Hz)	Na′	Pa′	Nb′	Na′-Pa′	Pa′-Nb′	Na′-Pa′	Pa′-Nb′	McPherson & Starr (1995)
0 (n=7)	24.9 (1.4)	33.0 (1.9)	47.5 (5.0)	0.73 (0.35)	0.85 (0.59)	35.8 (11.5)	48.3 (33.0)	N20 - 35.7 (7.0) P30 - 43.6 (18.5) N40 - 48.2 (9.1)
80 (n=10)	22.2 (3.1)	33.4 (3.5)	47.3 (5.8)	0.63 (0.27)	1.29 (0.54)	35.6 (11.5)	77.2 (29.9)	
160 (n=7)	23.9 (2.6)	36.0 (3.8)	51.0 (6.2)	0.87 (0.52)	0.84 (0.25)	40.2 (23.3)	48.8 (19.3)	
400 (n=7)	25.5 (2.7)	33.3 (3.3)	47.2 (5.7)	0.63 (0.24)	0.71 (0.45)	36.2 (11.1)	51.3 (26.2)	

exploring binaural interaction is that it is possible to determine the limit of binaural NON-fusion; paradigms involving IID or IPD do not permit assessment of this limen. Further work is needed, however, to explore the underlying mechanisms and the failure of the BIC to track perceptual changes, if not to search for an electrophysiological measure that is sensitive to binaural interaction under the IFD paradigm.

The various psychoacoustical results from this study can be shown to be similar to those in the literature or to differ in predictable ways (namely, for brief [versus long-duration] stimuli). The criteria for binaural nonfusion in the literature have been defined as the perception of two pitches (Thurlow and Bernstein, 1957; Perrott *et al.*, 1970), regardless of any change in the apparent location of either tone (Thurlow and Elfner, 1959). Thresholds based on these criteria were observed to fall around 9–30 Hz (estimated from figures in the articles cited), in good agreement with the 36-Hz threshold observed using the 250-ms tone bursts in the present study. The slightly larger value found here is expected from the more conservative criterion employed—clearly distinct pitches in the two ears, adopted in this study to guarantee complete loss of binaural fusion when the frequency difference reaches some limit. Indeed, the largest pair of ΔFs tested (i.e., 400 Hz) were found to be well beyond even the 209-Hz threshold observed using the briefer stimuli.

There are not many results of relevant comparison in the literature for the BIC of the AMLR. Nevertheless, it is worth noting that the percentage of binaural interaction to the summed monaural response observed in this study, roughly 40%, is similar to that observed by McPherson and Starr (1995) (the right side of Table I), despite their use of clicks rather than tone bursts. Taken together, these observations suggest another possible explanation for the lack of correlated changes in the BIC of the AMLR—the BIC of the AMLR is itself not frequency specific. Interestingly, DeVries and Decker (1988) also reported no frequency dependence for the binaural interaction of the ABR.

The BIC findings reported here, nevertheless, are not entirely inconsistent with behavioral results. Scharf (1969) found the overall loudness of a pair of equally loud tones presented dichotically with an interaural frequency difference to be independent of ΔF between ears. This is in contrast to the effects of monaural presentation of spectral components of increasing ΔF, particularly when the critical band is exceeded—loudness then grows with ΔF. This is in good agreement with our binaural response measures [Fig. 5(C)] and perhaps a clue to an interpretation of the results for the AMLR-BIC under the IFD paradigm. We assume that neurons giving rise to the mono-AMLR are largely contralateral to the cochlea of stimulation, yet not exclusively. Starting at the brainstem level, a minority of neurons contributing to the ascending activation of the human pathway, at least for simple onset responses, is expected to derive from ipsilateral activation (as, in effect, suggested by Moore, 1987). If this is true, the sum of the monaurally evoked responses inherently overestimates the true binaural response, on the assumption that the neurons that were ''minority'' partners in the monaural response are now a part of the ''majority'' populations on each side contributing to the binaurally evoked potential.

V. CONCLUSIONS

The BIC observed under the IFD paradigm fails to demonstrate systematic changes over a range of IFDs that cause clear and consistently detected changes in the perception of lateralization versus fusion, even with the relatively brief tone bursts required for AMLR measurement. Work is now needed to corroborate these findings and to pursue the basis(es) for the apparent lack of effect of an interaural difference otherwise expected to be effective considering effects of related interaural differences in time/phase, apropos the BIC literature. It is conceivable that the onset-response nature of the AMLR fails to reflect adequately the activity involved in the percept. For whatever reason, the BIC seems not to be frequency specific. Still, it is interesting that binaural loudness of sounds with IFD shows a similar lack of dependence on ΔF as observed for the AMLR-BIC. In this respect, the AMLR-BIC response does seem to reflect behavior. While the results were not as expected, the IFD paradigm remains useful in providing a measure of the limit of nonfu-

sion, i.e., of the ability to separate sounds of different frequencies between ears, reflecting the concept of the binaural critical band.

ACKNOWLEDGMENTS

This work derived from predissertation work in partial fulfillment of requirements for the Ph.D. degree. Gratitude is extended to Professor Connie Tompkins, University of Pittsburgh, for her help with the research design of this study, Professor Larry Feth, the Ohio State University, for his helpful comments in designing the psychoacoustical part of the study, and Lisa B. Taubman, M.A., University of Pittsburgh, for her assistance.

Ainslie, P. J., and Boston, J. R. (**1980**). ''Comparison of brainstem auditory evoked potentials for monaural and binaural stimuli,'' Electroencephalogr. Clin. Neurophysiol. **49**, 291–302.

Decker, T. N., and Howe, S. W. (**1981**). ''Auditory tract asymmetry in brainstem electrical responses during binaural stimulation,'' J. Acoust. Soc. Am. **69**, 1084–1090.

DeVries, S. M., and Decker, T. N. (**1988**). ''Frequency dependence of interear asymmetries and binaural interaction in the human ABR,'' Ear Hear. **9**(5), 275–282.

Dobie, R. A., and Norton, S. J. (**1980**). ''Binaural interaction in human auditory evoked potentials,'' Electroencephalogr. Clin. Neurophysiol. **49**, 303–313.

Durlach, N., and Colburn, H. (**1978**). ''Binaural phenomena,'' in *Handbook of Perception: Vol. 4, Hearing*, edited by Carterette and Friedman (Academic, New York), pp. 360–466.

Fowler, C. G. (**1989**). ''The bifrequency binaural interaction component of the auditory brainstem response,'' J. Speech Hear. Res. **32**, 767–772.

Furst, M., Levine, R. A., and McGaffigan, P. M. (**1985**). ''Click lateralization is related to the beta component of the dichotic brainstem auditory evoked potentials of human subjects,'' J. Acoust. Soc. Am. **78**, 1644–1651.

Hall, J. W. (**1992**). *Handbook of Auditory Evoked Responses* (Allyn and Bacon, Boston), Chap. 4, pp. 104–176.

Ito, S., Hoke, M., Pantev, C., and Lutkenhoner, B. (**1988**). ''Binaural interaction in brainstem auditory evoked potentials elicited by frequency-specific stimuli,'' Hear. Res. **35**, 9–20.

Levine, R. A. (**1981**). ''Binaural interaction in brainstem potentials of human subjects,'' Ann. Neurol. **9**, 384–393.

Mark, H. E., and Ratty, F. (**1990**). ''Frequency discrimination of single-, double-, and triple-cycle sinusoidal acoustic signals,'' J. Acoust. Soc. Am. **88**, 560–563.

McPherson, D. L., and Starr, A. (**1993**). ''Binaural interaction in auditory evoked potentials: Brainstem, middle-, and long-latency components,'' Hear. Res. **66**, 91–98.

McPherson, D. L., and Starr, A. (**1995**). ''Auditory time-intensity cues in the binaural interaction component of the auditory evoked potentials,'' Hear. Res. **89**, 162–171.

Moore, J. K. (**1987**). ''The human auditory brain stem as a generator of auditory evoked potentials,'' Hear. Res. **29**(1), 33–43.

Perrott, D. R., Briggs, R., and Perrott, S. (**1970**). ''Binaural fusion: Its limits as defined by signal duration and signal onset,'' J. Acoust. Soc. Am. **47**, 565–568.

Polyakov, A., and Pratt, H. (**1995**). ''Three-channel Lissajous' trajectory of the binaural interaction components of human auditory middle-latency evoked potentials,'' Hear. Res. **82**, 205–215.

Robinson, D. E. (**1971**). ''The effect of interaural signal-frequency disparity on signal detectability,'' J. Acoust. Soc. Am. **50**, 568–571.

Scharf, B. (**1969**). ''Dichotic summation of loudness,'' J. Acoust. Soc. Am. **45**, 1193–1205.

Scharf, B., Florentine, M., and Melselman, C. H. (**1976**). ''Critical band in auditory lateralization,'' Sens. Processes **1**, 109–126.

Spivak, L. G., and Seitz, M. R. (**1988**). ''Response asymmetry and binaural interaction in the auditory brain stem evoked response,'' Ear Hear. **9**(2), 57–69.

Thurlow, W. R., and Bernstein, S. (**1957**). ''Simultaneous two-tone pitch discrimination,'' J. Acoust. Soc. Am. **29**, 515–519.

Thurlow, W. R., and Elfner, L. F. (**1959**). ''Pure-tone cross-ear localization effects,'' J. Acoust. Soc. Am. **31**, 1606–1608.

Wrege, K. S., and Starr, A. (**1981**). ''Binaural interaction in human auditory brainstem evoked potentials,'' Arch. Neurol. **38**, 572–580.

Yama, M. F., and Small, A. M. (**1983**). ''Tonal masking and frequency selectivity for the monaural and binaural hearing systems,'' J. Acoust. Soc. Am. **73**, 285–291.

Pitch of amplitude-modulated irregular-rate stimuli in acoustic and electric hearing

Astrid van Wieringen[a)]
Lab. Exp. ORL, KULeuven, Kapucijnenvoer 33, 3000 Leuven, Belgium

Robert P. Carlyon and Christopher J. Long
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom

Jan Wouters
Lab. Exp. ORL, KULeuven, Kapucijnenvoer 33, 3000 Leuven, Belgium

(Received 30 August 2002; revised 4 April 2003; accepted 7 April 2003)

The pitch of stimuli was studied under conditions where place-of-excitation was held constant, and where pitch was therefore derived from ''purely temporal'' cues. In experiment 1, the acoustical and electrical pulse trains consisted of pulses whose amplitudes alternated between a high and a low value, and whose interpulse intervals alternated between 4 and 6 ms. The attenuated pulses occurred after the 4-ms intervals in condition A, and after the 6-ms intervals in condition B. For both normal-hearing subjects and cochlear implantees, the period of an isochronous pulse train equal in pitch to this ''4–6'' stimulus increased from near 6 ms at the smallest modulation depth to nearly 10 ms at the largest depth. Additionally, the modulated pulse trains in condition A were perceived as being lower in pitch than those in condition B. Data are interpreted in terms of increased refractoriness in condition A, where the larger pulses are more closely followed by the smaller ones than in condition B. Consistent with this conclusion, the A-B difference was reduced at longer interpulse intervals. These findings provide a measure of supra-threshold effects of refractoriness on pitch perception, and increase our understanding of coding of temporal information in cochlear implant speech processing schemes.
[DOI: 10.1121/1.1577551]

PACS numbers: 43.66.Ba, 43.66.Fe, 43.66.Hg [NFV]

I. INTRODUCTION

In electrical hearing, pitch can be modified either by manipulating the place of stimulation in the cochlea, or, independently, by varying the temporal pattern of stimulation at a fixed site. Understanding how both of these cues affect pitch perception is an important goal, which may yield benefits in improving speech processing strategies for cochlear implants. The present article extends the results of a recent study that examined ''purely temporal'' aspects of pitch perception (Carlyon *et al.*, 2002). As that previous article, our experiments with cochlear implant users were complemented by acoustic studies. These acoustic experiments used bandpass-filtered pulse trains containing no spectral cues for pitch (Shackleton and Carlyon, 1994; Carlyon, 1997), and which have been shown to produce analogous results to those obtained with electrical stimulation (McKay and Carlyon, 1999; Carlyon *et al.*, 2002). This general approach has two advantages. First, to the extent that the results with the two subject populations agree, it rules out interpretations that are specific to a particular mode of excitation. Second, because normal-hearing listeners are more abundant, experimenters can fine-tune their experimental design prior to studying the relatively scarce population of cochlear implant users.

In one of the experiments described by Carlyon *et al.* (2002), normal-hearing listeners and cochlear implant users compared the pitch of a ''4–6'' pulse train, in which the interpulse intervals alternated between 4 and 6 ms, to that of isochronous (i.e., fixed period) pulse trains having periods of between 3 and 8 ms. The point of subjective equality (PSE) corresponded to an isochronous pulse train having a period of 5.6 ms for the normal-hearing listeners and 5.7 ms for the cochlear implant users. The similarity between the acoustic and electric data reaffirmed previous findings (McKay and Carlyon, 1999) that temporal cues are perceived similarly by normal-hearing listeners and implantees. Moreover, the study of Carlyon *et al.* (2002) showed that the two alternative forced-choice task, requiring the subject to indicate the higher of two pitches, yielded the same pattern of results as the more time-consuming matching procedure.

Over the past few years, various studies have examined how modulation frequency, carrier rate, and modulation depth in periodic stimuli affect temporal pitch perception (e.g., McKay *et al.*, 1994, 1995; McKay and McDermott, 1996; McDermott and McKay, 1997; McKay and Carlyon, 1999). McKay *et al.* (1995) examined the effect of modulation depth by asking cochlear implantees to adjust the rate of an unmodulated pulse train so that its pitch matched that of a modulated pulse train containing a high-level pulse followed by one or more identical lower-level pulses. The rate of the unmodulated stimulus for which the modulated stimulus was judged equal in pitch varied systematically with modulation depth: at small modulation depths the matched rate was close

[a)]Electronic mail: astrid.vanwieringen@uz.kuleuven.ac.be

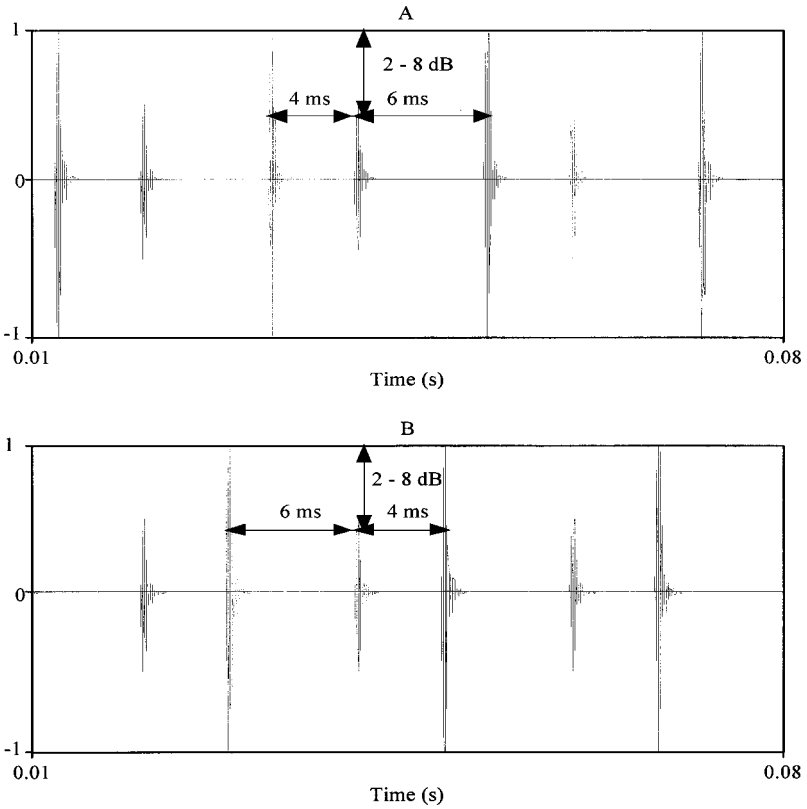


FIG. 1. Illustration of the "4–6 ms" acoustic pulse trains of conditions A (top panel) and B (bottom panel).

to the carrier rate, at larger modulation depths it approximated the modulation frequency. Further investigation with both implantees and normal-hearing subjects showed that pitch is reduced by increasing the modulation depth, and that the auditory system can derive two temporal patterns, one corresponding to the carrier rate and one corresponding to the modulation rate, from the *same* electrode, if the modulation rate is a submultiple of the carrier rate in the amplitude-modulated stimulus (McKay and Carlyon, 1999). However, when two equal-amplitude inharmonically related pulse trains are mixed within a single channel, subjects hear a single composite pitch, the value of which corresponds to that of the higher-rate pulse train (Carlyon, 1996; Carlyon *et al.*, 2002).

The extent to which normal-hearing listeners and cochlear implantees can derive temporal pitch information from amplitude-modulated pulse trains is further examined in the present study with irregular-rate stimuli, in which the interpulse intervals alternated between two values. In experiment 1, these values were 4 and 6 ms. The form of amplitude modulation we used was one in which every other pulse was attenuated by the same amount. In condition A, the attenuated pulses always occurred after the shorter (4-ms) interval [Fig. 1(a)], whereas in condition B they occurred after the longer (6-ms) interval [Fig. 1(b)]. While the "carrier rate" varied with varying interpulse interval, the modulation frequency remained the same (here: 10 ms). In experiment 2 longer periods were tested.

The experiments were motivated by two different hypotheses. One of these was based on our earlier finding (Carlyon *et al.*, 2002) that, when the interpulse interval in an unmodulated pulse train alternates between 4 and 6 ms, the perceived pitch is dominated by the longer of the two intervals. This leads to the prediction that reducing the length of the 6-ms interval should increase the perceived pitch, even if one increases the duration of the 4-ms interval by the same amount. We reasoned that such a pitch change should be determined not just by the physical timing of the pulses, but via differences in latency between the neural responses to high- and low-amplitude pulses. Carlyon *et al.* (2002) have shown that, for interpulse intervals up to about 12 ms, the pitch of an equal-amplitude pulse train is dominated by the longest first-order interval in the stimulus, and have argued that this "weighting" of longer intervals is at least partially due to mechanisms central to the auditory nerve (AN). If so, then increasing the longest first-order intervals in the AN response to a pulse train should reduce pitch in condition B, and cause it to be lower than in condition A, where a longer latency to the low-amplitude pulses would *decrease* the longest interpulse intervals (Fig. 1). Because the latency of AN responses to electrical pulses decreases with increasing amplitude (e.g., van den Honert and Stypulkowski, 1984; Javel and Shepherd, 2000), it is possible that such a pitch increase would occur in condition A [Fig. 1(a)], in which the longer interpulse intervals are ended by a higher-amplitude (and presumably shorter-latency) pulse. The latency hypothesis, therefore, would predict that pitch should be *higher* in condition A than in condition B. Latency data of the electrically evoked whole-nerve action potential (ECAP) measured in Ineraid cochlear implantees range between 0.38 and 0.41 ms (Brown and Abbas, 1990), while those of the acoustically evoked action potential (CAP) decrease from 4.1 to 1.2 ms nonlinearly as the intensity is increased from 10 to 90 dB SL (Yoshie, 1968). In that study acoustical latencies were 1.7 ms

TABLE I. Biographical data of the four LAURA implantees (CI1–CI4). The table includes their age, aetiology, approximate duration of profound deafness (>90 dB) in years (DD), number of implanted years (CI), vowel (SRV) and consonant (SRC) recognition scores in percentage correct, type of internal device (INT), channel under test (CH), current levels of threshold (THR), and most comfortable levels (MCL) for the 40 μs/phase biphasic pulse trains used in this study. Channel 1 is the most apical channel and channel 8 is the most basal one in the LAURA device.

Subject	Age	Aetiology	DD (years)	CI (years)	SRV (%)	SRC (%)	INT	CH no.	THR (μA)	MCL (μA)
CI1	20	Meningitis	3.5	7	53	45	5.6	2	250	600
CI2	44	Menière's	3	5	79	51	5.6	7	490	1000
CI3	33	Unknown	6	7	73	58	5.6	3	220	900
CI4	51	Congenital progressive	1	3	74	38	5.8	3	450	1050

at 60 dB SL, 1.5 ms at 70 dB SL, and 1.3 ms at 80 dB SL. An alternative hypothesis is suggested by the phenomenon of refractoriness, combined with the fact that the pitch of AM pulse trains decreases with increasing modulation depth. For stimuli such as ours it is probable that, at short interpulse intervals, the larger pulses will reduce the probability that AN fibers will fire to the subsequent smaller pulses—thereby transforming the neural response of the modulated pulse train into one more like the response of an unmodulated isochronous pulse train having a 10-ms period. This ''refractoriness'' effect will be stronger for condition A, where the smaller pulses follow the larger ones by only 4 ms, than in condition B, where the corresponding interval is 6 ms. According to this hypothesis, the pulse trains in condition A will be perceived as *lower* in pitch than those in condition B. A further prediction of this hypothesis is that any such refractoriness effects should decrease as the interpulse intervals increase. The two hypotheses were examined for different modulation depths and different rates in experiments 1 and 2.

In experiment 1 the acoustical and electrical amplitude-modulated irregular (''4–6 ms'') pulse trains (conditions A and B) were paired with unmodulated fixed-rate pulse trains in a two-alternative forced-choice (2-AFC) paradigm. A range of modulation depths was studied for both groups of listeners. The normal-hearing subjects also performed pitch matching. In experiment 2 temporal pitch perception in acoustical and electrical hearing was studied for longer interpulse intervals. Experiment 3 measured the detection thresholds for the attenuated pulses in a modulated pulse train.

II. GENERAL METHODS

A. Subjects

Four normal-hearing subjects (NH1–NH4) participated in the acoustical experiments. They had thresholds within 15 dB HL of the ISO-389 standard at all audiometric frequencies. Four postlingually deafened adults, three males and one female (CI1-CI4), took part in the electrical experiments. These subjects had been implanted with the LAURA device (Peeters *et al.*, 1993) in Belgium at the University ENT-Dept. of the St. Augustinus Hospital in Antwerp, or at the University Hospital in Leuven. Details concerning their aetiology, duration of profound deafness, age at implantation, number of implanted years, speech perception results, version of internal device implanted, and channel under test are given in Table I. They had also participated in similar pitch experiments with unmodulated pulse trains (Carlyon *et al.*, 2002). Based on closed set vowel and consonants tests (SRV and SRC, respectively, Table I), they are considered moderate to good performers (van Wieringen and Wouters, 1999).

B. LAURA device

The LAURA device consists of eight electrode pairs, each forming a bipolar channel. The implanted components receive input from the external speech processor via a transcutaneous inductive link. Amplitude is coded by changing the current level of each electrical pulse.

Two different types of electrode configurations are in use: types 5.6 and 5.8. On both arrays the electrodes are placed on the most apical 17 mm of a 25-mm silicon carrier with a slightly curved 3-mm tip. Type 5.8, the most recent design, mainly differs from type 5.6 in the orientation of the electrode contacts within the channels. It was changed from offset-longitudinal (type 5.6) to longitudinal (type 5.8), and the distance between pairs of electrodes, also referred to as *channels*, was increased from 2 to 3 mm. Another difference between the two designs is the diameters of the ball contacts, which are 0.24 and 0.34 mm for types 5.6 and 5.8, respectively. Only one of the four subjects in this study (CI4) uses type 5.8.

C. Stimuli and procedure

The acoustical stimuli were pulse trains having an overall duration of 400 ms, generated in MATLAB 5.1 at a sampling rate of 44 100 samples/s and filtered digitally between 3900 and 5300 Hz to remove any frequency components that are resolvable by the peripheral auditory system. The digital filter approximated an analog eighth-order Butterworth filter. The attenuation at the cutoff frequencies was 3 dB relative to the level in the center of the passband, and the attenuation at half an octave below the lower cutoff and at half an octave above the upper cutoff was at least 24 dB relative to that at the cutoff frequencies. All stimuli were gated with 50-ms onset and offset cosine ramps. They were presented at 54 dB SPL against a background of continuous pink noise. The pink noise had a spectrum level of 9.5 dB SPL at 4 kHz.

The acoustical stimuli were routed via a portable PC (Toshiba, Satellite Pro 420 CDT) and a 16-bit PCMCIA card (WAVjammer, New Media Corp) to a mixer (Eurorack

TABLE II. Top: pitch matches of each normal-hearing subject (NH1–NH4). Each pitch match is the geometrical mean of 18 estimates. Geometrical standard errors (S.E.) are included in the table. Bottom: averaged modulation depths (in percentage of dynamic range) and standard deviations (S.D.) of each cochlear implant subject (CI1–CI4). These modulation depths are perceptually equivalent to 2, 4, and 6 dB in normal hearing. See Sec. II B in text for more details.

	2 dB (ms)	S.E. (ms)	4 dB (ms)	S.E. (ms)	6 dB (ms)	S.E. (ms)
NH1	5.6	0.3	7.8	0.06	7.2	0.1
NH2	5.2	0.2	6.1	0.1	8.2	0.2
NH3	5.5	0.5	6.1	0.2	7.8	0.6
NH4	5.2	0.3	6.3	0.5	8.7	0.6
Gr. Mean	**5.4**		**6.6**		**7.9**	
	Modulation depths expressed in percent dynamic range					
	2 dB	S.D. (%)	4 dB	S.D. (%)	6 dB	S.D. (%)
CI1	3%	1.1	6%	0.8	12%	2.8
CI2	9%	2.7	35%	1.6	51%	1.0
CI3	7%	1.2	9%	2.0	20%	8.5
CI4	4%	0.4	8%	1.6	13%	0.7

MX1604A), where they were combined with the pink noise from a CD player (Sony, CDP-209) and played to the right earpiece of a headset (TDH39).

The electrical stimuli were 400-ms biphasic current pulse trains (40 μs/phase), presented to a single pair of electrodes. They were sent to the electrode array via the DSP-TIC30 board in the (portable) computer. Thresholds and most comfortable levels (MCLs) were determined prior to testing with a 5-ms biphasic pulse train. Current levels (in μA) are listed in Table I. For experiments 1 and 2 the stimulus amplitudes were mapped between threshold and MCL of the subject's channel, and are expressed as a percentage of this dynamic range (in terms of μA). For both electrical and acoustical experiments, stimulus presentation and response collection were controlled with a hardware/software interface (APEX, Geurts and Wouters, 2000).

Figure 1 illustrates the acoustical amplitude modulated pulse trains in which the interpulse interval alternated between 4 and 6 ms (''4–6 ms'' stimuli) in conditions A and B of experiment 1. These stimuli were compared with or matched to isochronous stimuli with a range of repetition rates. For more details see separate experiments.

III. EXPERIMENT 1

A. Overview

The pitch of amplitude-modulated pulse trains in both normal-hearing listeners and cochlear implantees was examined in experiment 1. In the main part of the experiment, modulated, irregular, pulse trains (''4–6 ms'') were paired, on a trial-by-trial basis, with unmodulated, isochronous pulse trains having periods between 4 and 10 ms. On each trial, subjects judged which of the two pulse trains had the higher pitch. The PSE was estimated from a function relating the period of the unmodulated isochronous stimuli to the proportion of times that stimulus was judged higher in pitch than the modulated ''4–6 ms'' pulse train. This method proved appropriate in previous pitch matching experiments (McKay *et al.*, 1994; Carlyon *et al.*, 2002). It has the advantages of being fast, and of helping subjects to focus on pitch and to ignore quality differences between the irregular stimuli and the comparison isochronous stimuli that were presented in random order within an experimental run (Carlyon *et al.*, 2002). For the normal-hearing listeners, this exercise was carried out at modulation depths of 2, 4, 6, and 8 dB. Choosing modulation depths for implant users that correspond to these acoustic values is not straightforward, and so we performed a preliminary experiment, as described below. The normal-hearing subjects also matched the pitches of the irregular-rate amplitude-modulated standard stimuli (A and B) to unmodulated signals in an adjustment task.

B. Method

1. Preliminary experiment

First, four normal-hearing subjects were asked to adjust the rate of an unmodulated stimulus so that its pitch matched that of a 5-ms stimulus modulated by 2, 4, or 6 dB, depending on the conditions. The rates of the unmodulated stimuli ranged from 4.5 to 12.5 ms in steps of 4%. Subjects had to choose the higher pitch stimulus in a 2AFC, adaptive procedure, and the (one-down, one-up) decision rule targeted the matching value at 50%. The interstimulus interval was 500 ms. Data are listed in Table II. The resulting pulse periods of 5.4 ms (2 dB), 6.6 ms (4 dB), and 7.9 ms (6 dB), each based on 18 estimates, decreased in rate with increasing modulation depth—consistent with previous results (McKay *et al.*, 1994; McKay and Carlyon, 1999). Then, each implant user adjusted the modulation depth of a 5-ms pulse train so that its pitch matched each of the three pulse periods (5.4, 6.6, and 7.9 ms). The modulation depths of the 5-ms pulse trains were changed in steps of 2% between 80% and 100% of the dynamic range, and in steps of 5% below 80%. Once again, the subjects were required to choose the higher pitch stimulus in a 2AFC task using a one-down one-up decision rule. The stimuli were mapped between the threshold and most comfortable level of the implantee (in terms of μA). The larger pulses were presented at 100% (=MCL), while the smaller pulses were presented at 100% minus the modulation depth. No feedback was provided. The experiment showed

that when the cochlear implant users matched to the mean rates of the normal-hearing subjects, the modulation depth of the modulated stimulus was greatest when the pitch of the unmodulated stimulus had the lowest rate (7.9 ms, bottom part of Table II).

The resulting modulation depths, averages of three estimates, were then used in the main part of experiment 1, and are referred to as ''perceptually equivalent'' to 2, 4, and 6 dB acoustic modulation depths.

2. Forced-choice pitch judgments

On each trial, one modulated, irregular, stimulus (''4–6 ms,'' A and B) was paired in random order with one of five unmodulated isochronous signals, having periods of 4, 5.5, 7, 8.5, and 10 ms. This range of signals was chosen to range between stimuli consistently sounding lower in pitch to stimuli consistently sounding higher in pitch than the modulated irregular stimuli (on the basis of pilot experiments with normal-hearing listeners). Because amplitude modulation is known to reduce the perceived pitch both in electric and acoustic hearing (McKay and Carlyon, 1999) the range falls largely above the PSE of 5.6 ms for unmodulated stimuli (Carlyon *et al.*, 2002). For each implant subject all stimuli were presented at the same current level, since a pilot loudness balancing test had shown that loudness did not strongly vary with repetition rate over the rate range used in the present experiment (with periods of 4–8 ms). This is consistent with data of McKay and McDermott (1998) who showed that differences in loudness across conditions were not significant when each condition was adjusted independently to its comfortable loudness level. Moreover, the present study is mainly concerned with the relative comparison of pitch between conditions A and B. In both conditions the current levels were the same.

The test consisted of a block of 50 trials containing stimuli of one modulation depth (i.e., the modulated, irregular, signal) paired ten times with each of the five isochronous stimuli in random order (either in the first interval or in the second interval). The interstimulus time interval was 500 ms. The subject indicated which interval was higher in pitch, after which the next pair was presented. No feedback was provided. Each block of 50 trials was randomized independently. During each testing session, subjects listened to five blocks of trials with the condition A stimulus modulated at a certain depth, and five blocks of trials with the condition B stimulus modulated by the same amount. Eight different stimulus conditions were completed by the normal-hearing subjects (4 modulation depths$\times$A and B), and six by the cochlear implantees (3 modulation depths$\times$A and B). These conditions were randomized over the different testing sessions. Testing sessions lasted about 25 min. The normal-hearing subjects did 25 blocks of 50 trials in conditions A and B for each modulation depth (a total of 1250 estimates in each condition), and the cochlear implant subjects did 10 blocks of 50 trials (a total of 500 estimates) in each condition.

3. Pitch matching

The four normal-hearing subjects also performed a pitch matching experiment in which the standard stimuli were the irregular-rate pulse trains (''4–6 ms,'' A and B), modulated at 2, 4, 6, and 8 dB. Although we have argued that this method is less efficient than the forced-choice paradigm (Carlyon *et al.*, 2002), it allows one to check that the pitches heard for a given stimulus are distributed unimodally across trials, and is a technique that has been more commonly used in research on pitch perception. The irregular-rate pulse trains were matched against a set of 29 unmodulated trains with periods varying from 2.1 to 14 ms in steps of 7%, and bandpass filtered between 3900 and 5300 Hz. These stimuli were identical to those used by Carlyon *et al.* (2002). During the matching test, subjects continuously heard a triad of stimuli, the amplitude-modulated irregular pulse train, an isochronous pulse train, and the amplitude-modulated irregular pulse train again. Interstimulus time was 400 ms within, and 1500 ms between, triads. Subjects were required to adjust the pitch of the second pulse train in order to match to the pitch of the first and third stimuli, by clicking ''---,'' ''--'' ''-'' for a lower rate or ''+,'' ''++,'' ''+++'' for a higher rate to cause one of three different step size changes in rate. There were six different starting values of the isochronous pulse trains. When subjects were satisfied with the match they could click ''OK,'' after which the next triad was presented. The four normal-hearing subjects performed thirty matches in each of the different stimulus conditions (4 modulation depths$\times$A and B). All conditions were tested separately.

C. Results

1. Forced-choice pitch judgments: Normal-hearing listeners

Psychometric functions of the normal-hearing subjects, illustrating the proportion of trials in which the five different isochronous stimuli are judged higher in pitch than the irregular-rate pulse trains (condition A, unfilled symbols, and condition B, filled symbols) are shown for the four different modulation depths in Fig. 2. The crossover points at 50%, i.e., the points of subjective equality, were estimated by a probit function (SPSS 10.0, 2000). The individual values of each condition and subject are listed in Table III, together with the geometric means and standard errors.

Two main findings are worthy of comment. First, as expected, the PSE move to larger values as modulation depth increases, consistent with the results of McKay and Carlyon (1999), and also consistent with the present data in comparison to those of the unmodulated pulse trains (5.6 ms) by Carlyon *et al.* (2002). Second, the PSE in condition A tends to occur at longer periods (lower pitches) than in condition B. This can be seen in Table III, and by the fact that the lines connecting the open symbols in Fig. 2 tend to fall to the right of those connecting the filled symbols. Both of these trends were confirmed by a two-way (modulation depth$\times$stimulus type) repeated-measures ANOVA, which revealed significant effects of modulation depth [$F(3,9)=95.1$, $p<0.001$] and stimulus type [$F(1,3)=21.1$, $p<0.02$]. However, although

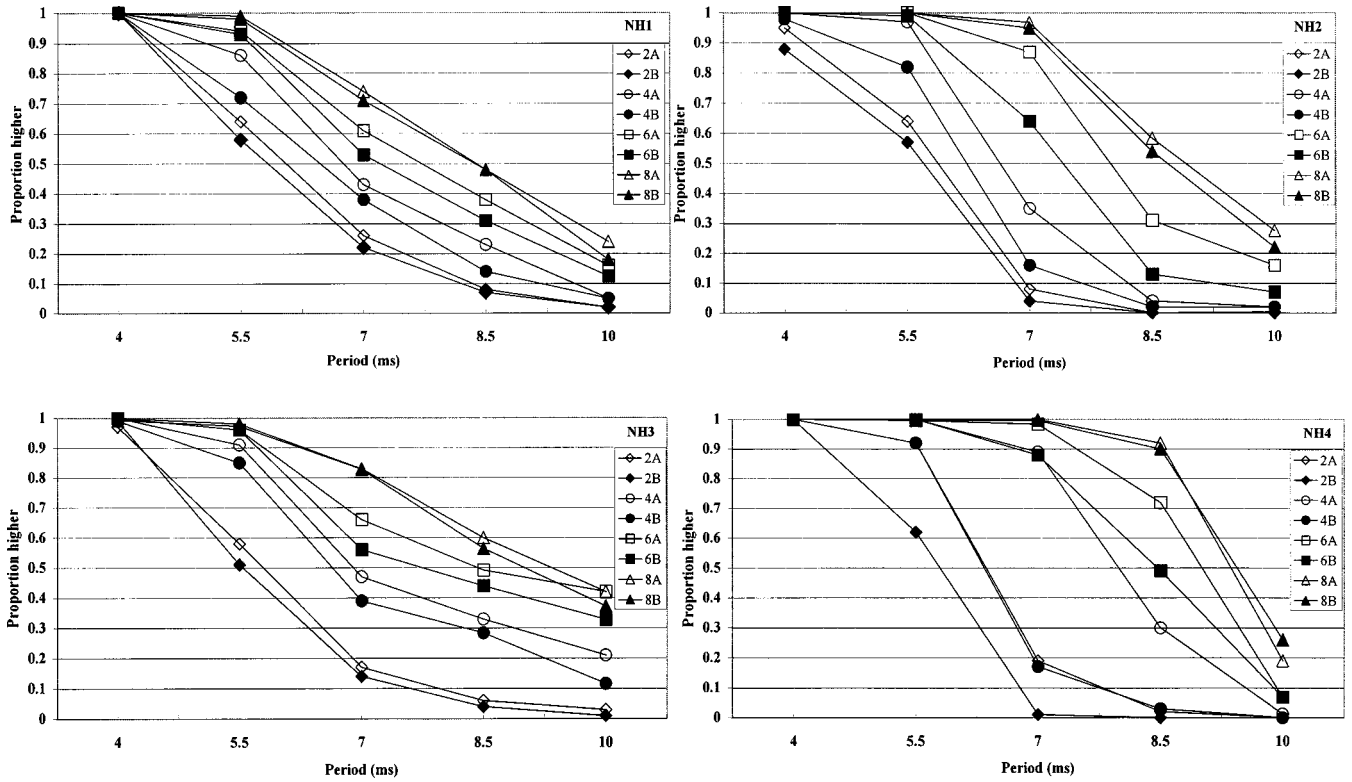


FIG. 2. The proportion of trials in which the unmodulated isochronous stimuli were judged as higher in pitch than the amplitude-modulated irregular-rate stimuli (A and B). Each panel shows an individual normal-hearing subject's responses. The plots of the four normal-hearing subjects each display results of 4 modulation depths (2, 4, 6, and 8 dB) in condition A (unfilled symbols) and in condition B (filled symbols).

there is some indication that the difference between conditions A and B decreased at the largest modulation depth, this interaction failed to reach significance [$F(3,9)=2.8$, $p=0.1$]. The extent to which A and B differ from each other can also be analyzed by fitting the two psychometric functions (of each subject and condition separately) with a probit function. With the exception of NH1 at 2 dB, and NH1 and NH3 at 8 dB (the effect of increasing modulation depth), all conditions are significant (Table III).

The fact that the modulated stimuli in condition A had

TABLE III. Period of an acoustical isochronous pulse train (in ms) judged equal in pitch to each of the acoustical amplitude-modulated irregular-rate ("4–6 ms") pulse trains in conditions A and B. Means and standard errors are geometric. The χ^2 statistic describes the extent to which the PSEs of A and B of the normal-hearing subjects differ from each other; χ^2 values in bold are significant (see text). The value of χ^2 statistic (df=1) needs to exceed 3.84 to be significant at $p<0.05$ and 6.64 to be significant at $p<0.01$.

	2 dB—A (ms) intercept	2 dB—B (ms) intercept	χ^2	4 dB—A (ms) intercept	4 dB—B (ms) intercept	χ^2
NH1	6.3	6.1	2.6	7.1	6.7	**15.7**
NH2	5.8	5.5	**20.1**	6.9	6.3	**45.6**
NH3	6.0	5.8	**4.3**	7.7	7.2	**16.6**
NH4	6.5	5.7	**123.3**	8.1	6.5	**386.6**
Mean	**6.2**	**5.8**		**7.4**	**6.7**	
S.E.	**0.3**	**0.3**		**0.6**	**0.4**	
	6 dB—A (ms)	6 dB—B (ms)	χ^2	8 dB—A (ms)	8 dB—B (ms)	χ^2
NH1	8.0	7.6	**7.4**	8.5	8.3	2.2
NH2	8.3	7.5	**68.0**	9.0	8.7	**4.0**
NH3	8.9	8.4	**9.0**	9.3	9.1	2.5
NH4	8.8	8.4	**31.1**	9.4	9.5	**158.3**
Mean	**8.5**	**8.0**		**9.0**	**8.9**	
S.E.	**0.5**	**0.5**		**0.4**	**0.5**	

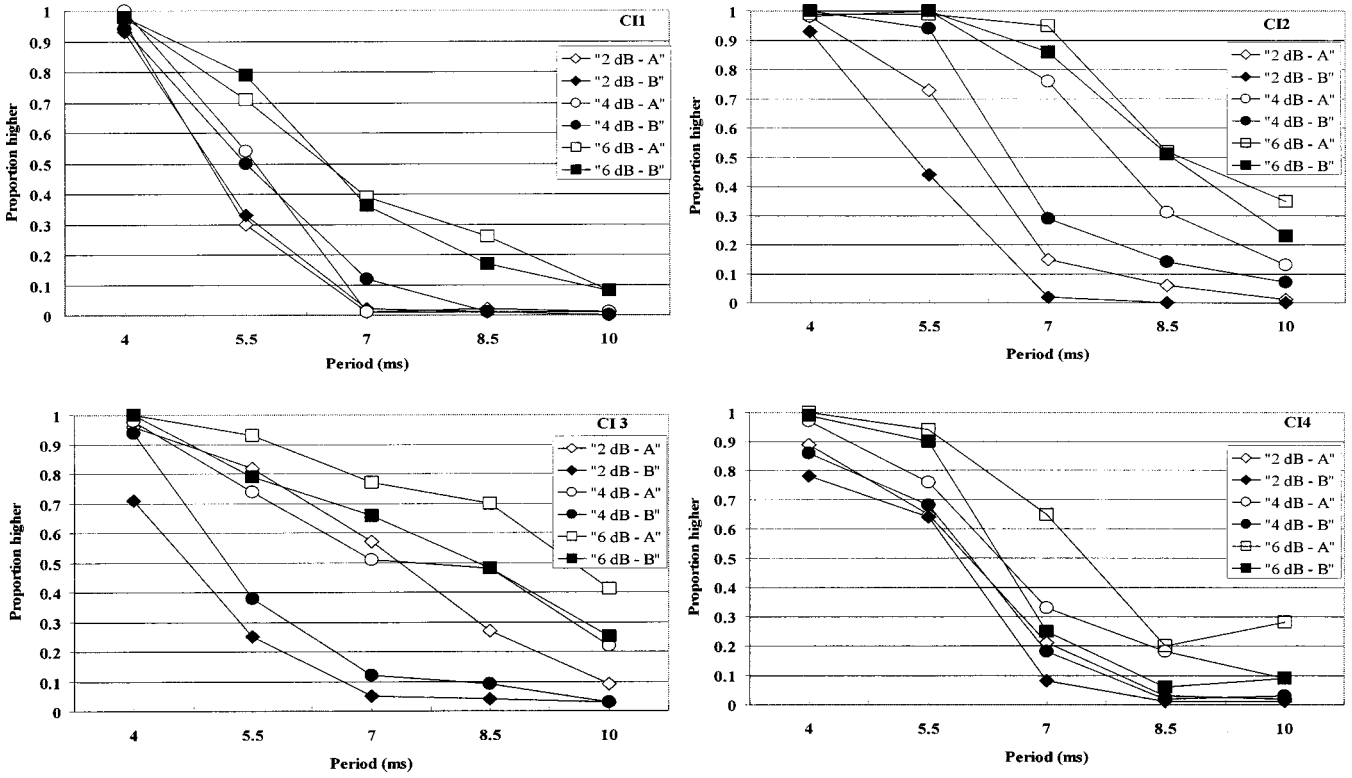


FIG. 3. Proportion of trials in which the unmodulated isochronous stimuli were judged as higher in pitch than the amplitude-modulated irregular-rate stimuli (A and B). The plots of the four cochlear implant subjects each include six stimulus conditions, i.e., 3 modulation depth×condition A (unfilled symbols) and B (filled symbols).

lower pitches than the stimuli in condition B is consistent with the idea that the larger pulses reduce the probability of neural responses to the smaller ones in condition A, thereby reducing the effective modulation depth and, hence, the pitch. This refractoriness hypothesis is studied further with longer periods in experiment 2.

2. Forced-choice pitch judgments: Cochlear implant users

The psychometric functions of the cochlear implant users, illustrating the proportion of trials in which the isochronous pulse trains were judged higher in pitch than the two irregular-rate pulse trains, are shown in Fig. 3. Probit esti-

TABLE IV. Period of an electrical isochronous pulse train (in ms) that is subjectively equal in pitch to electrical amplitude-modulated irregular-rate (''4–6 ms'') pulse trains in conditions A and B. Means and standard errors are geometric. Modulation depths are perceptually equated to 2, 4, and 6 dB. The χ^2 statistic describes the extent to which the PSEs of A and B of the cochlear implantees differ from each other; χ^2 values in bold are significant (see text).

	2 dB—A (ms) intercept	2 dB—B (ms) intercept	χ^2	4 dB—A (ms) intercept	4 dB—B (ms) intercept	χ^2
CI1	5.2	5.2	0.2	5.7	5.6	0.01
CI2	6.2	5.3	**39.1**	8.1	7.0	**44.0**
CI3	7.3	4.6	**176.3**	7.8	5.6	**115.3**
CI4	5.9	5.5	**8.1**	6.8	5.9	**27.1**
Mean	**6.1**	**5.1**		**7.0**	**6.0**	
S.E.	**0.9**	**0.4**		**1.2**	**0.7**	

	6 dB—A (ms)	6 dB—B (ms)	χ^2
CI1	6.9	6.9	0.11
CI2	9.1	8.7	3.3
CI3	9.5	7.9	**25.6**
CI4	8.0	6.8	**47.3**
Mean	**8.3**	**7.5**	
S.E.	**1.3**	**0.9**	

TABLE V. Geometrical means and standard errors of the 30 pitch matches made by the four normal-hearing listeners for each of the four modulation depths and two conditions (A and B).

	2 dB—A (ms)	2 dB—B (ms)	4 dB—A (ms)	4 dB—B (ms)	6 dB—A (ms)	6 dB—B (ms)	8 dB—A (ms)	8 dB—B (ms)
NH1	6.0	5.8	6.5	6.2	7.0	6.5	7.2	7.3
NH2	6.7	6.7	7.3	7.2	7.9	8.0	8.7	8.7
NH3	6.0	5.8	6.4	6.6	7.4	7.5	8.1	8.5
NH4	5.2	5.2	7.9	6.0	8.1	7.1	9.0	9.1
Mean	**5.9**	**5.9**	**7.0**	**6.5**	**7.6**	**7.3**	**8.2**	**8.4**
S.E.	**0.7**	**0.6**	**0.8**	**0.5**	**0.5**	**0.6**	**0.8**	**0.8**

mates reflecting the point of subjective equality are listed in Table IV.

The two main findings observed with the normal-hearing listeners are also apparent in the data for the cochlear implant users. That is, the PSEs correspond to lower pitches (longer periods) at larger modulation depths, and PSEs in condition A tend to occur at longer periods than in condition B (Fig. 3). These findings were confirmed for three of the four CI subjects by fitting the two psychometric functions of each subject and condition separately. The pitch difference between A and B was statistically significant, for CI3 and CI4 at all modulation depths and for CI2 at the two smaller modulation depths. Only CI1 perceived a similar pitch in conditions A and B. Hence, as with the normal-hearing listeners, the data obtained with cochlear implant users are consistent with a refractoriness hypothesis, and where the difference in PSE between conditions A and B reflects the difference in the size of the refractory effect between the two conditions. As the intervals between the large and small pulses become larger, the refractory effects in both conditions should tend towards zero, and the difference in PSE between the two conditions should eventually disappear. This hypothesis was examined further in experiment 2 with both normal-hearing subjects and cochlear implantees.

3. Pitch matching

Table V lists the geometrical means of the 30 matches of each normal-hearing subject at each modulation depth and condition in the pitch matching experiment. Consistent with the results of the forced-choice measurements, the point of subjective equality shifted to progressively longer periods as modulation depth was increased. When the data of the normal-hearing subjects were subjected to a three-way (paradigm$\times$condition$\times$modulation depth) ANOVA, the only factor to reach significance was an overall effect of modulation depth [$F(3,9)=53.9$, $p<0.01$]. It should be noted, however, that, when the results of the pitch-matching experiment were subjected to a separate ANOVA, there was not a significant difference in pitch between conditions A and B. This may have been due to a lack of power: only 30 matches were obtained per subject and condition compared to 1250 pitch matches in the forced-choice paradigm. Furthermore, as pointed out by Carlyon *et al.* (2002), at the end of a pitch match the pitches of the two stimuli are, by definition, very close, and subjects may then be distracted by the difference in quality between modulated and unmodulated sounds. Certainly, the absence of a significant condition $\times$paradigm interaction cautions against overinterpreting this difference.

Two further aspects of the pitch matching results, not described in the tables, are worth mentioning. First, the distributions of matches for a given subject and condition were unimodal, and, second, there were no matches to 4 or 10 ms. In a previous paper (Carlyon *et al.*, 2002), we obtained a similar finding for pitch matches to unmodulated irregular pulse trains. That article pointed out that the pattern of matches was inconsistent with the predictions of the autocorrelation models of pitch proposed by Meddis and Hewitt (1991), according to which subjects should match to 10 ms (the actual period of the stimulus), with a ''second choice'' of 4 ms (the shorter of the two first-order intervals). The present findings confirm this conclusion, and show that, even with a modulation depth as large as 8 dB, subjects do not match to the 10 ms, nor to the 5 ms predicted by the revised model of Meddis and O'Mard (1997).

IV. EXPERIMENT 2: PITCH PERCEPTION WITH LONGER INTER-PULSE INTERVALS

A. Method

Four normal-hearing listeners performed a 2AFC pitch comparison task, where the subject had to judge whether an amplitude-modulated irregular-rate pulse train (A or B) was higher or lower in pitch than an unmodulated isochronous pulse train. Stimuli were presented with the interpulse intervals alternating between 8 and 12 ms (''8–12 ms'') and the periods of isochronous pulse trains at 8, 10, 12, 14, and 16 ms; with the interpulse intervals alternating between 12 and 18 ms (''12–18 ms'') and the periods of isochronous pulse trains at 12, 15, 18, 21, and 24 ms; with the interpulse intervals alternating between 16 ms and 24 ms (''16–24 ms'') and the periods of isochronous pulse trains at 16, 20, 24, 28, and 32 ms; and with the interpulse intervals alternating between 24 and 36 ms (''24–36 ms'') and the periods of isochronous pulse trains at 24, 30, 36, 42, and 48 ms.

Only one modulation depth, 2 dB, was studied because this was the depth which had yielded the most significant difference between conditions A and B for both normal-hearing and cochlear implant subjects in experiment 1. Rate discrimination thresholds increase markedly at very low overall rates, but good performance can be preserved down to lower overall rates by filtering the stimuli into lower frequency regions (Krumbholz *et al.*, 2000). Thus, we chose to

TABLE VI. Period of an acoustical isochronous pulse train (in ms) that is subjectively equal in pitch to the acoustical 2-dB amplitude-modulated irregular-rate pulse trains in condition A and B. The χ^2 statistic describes the extent to which PSEs of A and B of the normal-hearing subjects differ from each other. None of these values of χ^2 are statistically significant. Means and standard errors show no significant differences at these lower rates.

	8–12 ms—A (ms) intercept	8–12 ms—B (ms) intercept	χ^2	12–18 ms—A (ms) intercept	12–18 ms—B (ms) intercept	χ^2
NH1	12.0	11.7	1.4	17.7	17.4	0.4
NH2	11.4	12.4	1.4	17.6	17.9	0.5
NH3	11.8	11.7	0.1	17.9	18.0	0.2
NH4	11.5	11.2	2.1	17.1	17.0	0.04
Mean	**11.7**	**11.7**		**17.6**	**17.6**	
S.E.	**0.3**	**0.5**		**0.4**	**0.5**	
	16–24ms—A (ms)	16–24ms—B (ms)	χ^2	24–36 ms—A (ms)	24–36 ms—B (ms)	χ^2
NH1	23.3	22.4	2.3	36.7	36.7	0.03
NH2	23.3	23.4	0.2	37.8	37.3	0.08
NH3	23.2	23.3	0.05	35.6	35.8	0.3
NH4	22.5	22.1	0.5	37.1	36.1	2.2
Mean	**23.1**	**22.8**		**36.8**	**36.5**	
S.E.	**0.4**	**0.7**		**0.9**	**0.7**	

filter our pulse trains between 1375 and 2875 Hz in order to maximize the accuracy of listeners' pitch judgments, while ensuring that even with the shortest period (12 ms) studied, no harmonics of the pulse train were resolvable by the peripheral auditory system (Shackleton and Carlyon, 1994). Even at very low rates the acoustical and electrical pulse trains elicit a clear pitch. Studies with normal-hearing listeners have shown that the lowest limit of melodic pitch varies with the specific stimuli and paradigm used but is approximately 20 Hz. For cochlear implantees, the limit is around 50 Hz (Zeng, 2002).

Cochlear implant users performed a subset of the conditions carried out by the normal-hearing listeners, namely the ''8–12 ms'' stimulus modulated by the perceptual equivalent of 2 dB. The electrical modulation depths used were taken from the preliminary experiment. Previously, McKay and Carlyon (1999) showed that the effect of modulation depth on pitch matches did not vary markedly as overall rate was varied between 140 pps (7.14 ms) and 212 pps (4.7 ms). However, it should be acknowledged that the present experiment extends to lower overall rates than were examined in that study. Data of both normal-hearing subjects and cochlear implantees are based on 100 trials per point (for a total of 500 estimates per subject).

B. Results

1. Normal-hearing subjects

The individual and averaged crossover points, estimated with probit analyses, are listed in Table VI. As in the higher rate conditions (Table III), the point of subjective equality is close to, but slightly less than, the longest first-order interval in the stimulus (e.g., 12 ms in the ''8–12 ms'' stimulus). However, contrary to the ''4–6 ms'' condition, there is no consistent difference between conditions A and B. This follows from a prediction of the refractoriness hypothesis: the refractory effects in conditions A and B should tend to zero at longer overall periods, and the difference between the matches obtained with the two stimuli should also decrease.

At lower rates, pitch in conditions A and B are perceived similarly. This is also confirmed by the nonsignificant χ^2 statistics, derived by fitting the A and B psychometric functions in each condition separately (Table VI). Our experiments show that an effect of refractoriness on pitch perception could only be observed with the relatively short (4 and 6 ms) interpulse intervals of experiment 1.

2. Cochlear implant users

Crossover points of the ''8–12 ms'' condition, estimated from probit analyses, are listed in Table VII. At this lower rate there is a hint that PSEs in condition A occur at longer periods (mean = 10.0 ms) than in condition B (9.4 ms). However, this χ^2 statistic on the difference between the psychometric functions of A and B was only significant for listener CI2. Also, the pitch matches of implantees were, overall, at lower intervals than of normal-hearing subjects in the same condition. Two *t*-tests on the crossover points showed that

TABLE VII. Period of an electrical isochronous pulse train (in ms) judged equal in pitch to the acoustical amplitude-modulated (''8–12 ms'') irregular-rate pulse trains in conditions A and B. The modulation depth of every other pulse is ''2 dB.'' Means and standard errors are geometric. The χ^2 statistic describes the extent to which the PSEs of A and B of the cochlear implantees differ from each other. Only data of CI2 are significant. (See text for more details.)

	8–12 ms—A (ms) intercept	8–12 ms—B (ms) intercept	χ^2
CI1	10.5	10.0	8.1
CI2	10.7	9.2	**54.0**
CI3	8.7	8.5	0.3
CI4	10.1	9.9	0.4
Mean	**10.0**	**9.4**	
S.E.	**1.0**	**0.7**	

the perceptual differences between the normal-hearing subjects and cochlear implantees (Table VI, ''8–12 ms,'' and Table VII) were significant in terms of both A ($p<0.05$) and B ($p<0.002$) conditions.

V. EXPERIMENT 3: DETECTION THRESHOLDS OF ATTENUATED PULSES IN MODULATED PULSE TRAINS

A. Rationale

Experiments 1 and 2 showed that the pitch of amplitude-modulated pulse trains decreased with increasing modulation depth for both normal-hearing and implanted subjects. The results of these experiments were interpreted in terms of the larger pulses in each train reducing the probability of neurons responding to the smaller pulses in each pulse train. The first part of experiment 1 showed that, even at the largest modulation depth studied, the pitch of a 5-ms pulse train had decreased substantially (to 7.9 ms), but not as far as the 10 ms that would result from an infinite modulation depth. Hence the results of our first two experiments may be interpreted as a *supra-threshold* measure of refractoriness, with differences in refractory effects between conditions A and B being reflected in the difference between their pitches. In experiment 3, we measured the modulation depth at which subjects could just discriminate between a modulated stimulus at a given interpulse interval (IPI_M) and an unmodulated stimulus with a period of $2\times IPI_M$. This provided a complementary measure of refractoriness, one based on detection thresholds for the smaller pulses, and allowed us to address two questions. First, how far above detection threshold were the smaller pulses in experiment 1? Second, do the intersubject differences observed in that experiment correlate with those observed at detection threshold?

B. Method

Three separate ranges of bandpass filtered acoustical stimuli and electrical pulse trains were generated. In each range, the interpulse interval was either 4, 5, or 6 ms. Every other pulse was decreased in steps of either 1 dB for the normal-hearing subjects or by 2% for the cochlear implantees. These modulated stimuli were compared to an unmodulated lower rate reference stimulus with periods of 8, 10, and 12 ms, respectively.

Detection thresholds for the lower-amplitude pulses were determined using a three-interval two AFC task, in which the unmodulated (lower-rate) reference stimulus was administered in the first interval and in one other interval, while the (higher-rate) amplitude-modulated signal was presented in interval two or three. Subjects were required to indicate which of the latter two intervals contained the ''odd'' stimulus. Feedback was provided. Thresholds were determined in an adaptive procedure using a two-down, one-up decision rule (Levitt, 1971). Testing was terminated after 12 reversals, the last six of which were used to determine the threshold. This value (in dB or %) is the level at which the amplitude-modulated (higher rate) stimulus is indistinguishable from the unmodulated lower rate one at a 71% criterion level. For each stimulus condition six threshold measures were obtained of the normal-hearing subjects and at least three threshold measures were obtained of the cochlear implantees.

C. Results

Figure 4 illustrates the average detection thresholds for the lower-amplitude pulses at three pulse rates in the normal-hearing subjects (left) and cochlear implantees (right). Thresholds indicate the value at which the modulated stimulus cannot be distinguished from the unmodulated stimulus, i.e., the point at which the larger pulses are masking the smaller ones. Standard deviations, averaged over the three different rates, are relatively small, i.e., 2.4, 2.1, 1.9, and 2.7 dB for NH1–NH4, respectively, and 8.5%, 4.0%, 2.4%, and 4% for CI1–CI4, respectively. The results show that all the normal-hearing listeners could detect the smaller pulses at modulation depths greater than the largest depth (8 dB) used in experiment 1. Similarly, the thresholds of the implantees corresponded to modulation depths greater than the greatest values used in that experiment; these values, which corresponded to a depth perceptually equivalent to 6 dB (Table II), are replotted on the right-hand side of the figure. This illustrates that the smaller pulses in experiment 1 were not completely masked by the larger ones, either for electrical or acoustical stimulation.

For the cochlear implant users, there is evidence that these detection thresholds correlate with the effects of refractoriness on pitch judgements in experiment 1. In the first part of that experiment, listener CI2 required 5-ms pulse trains with relatively large modulation depths (i.e., 9%, 35%, and 51%, Table II) to match the pitch of unmodulated pulse trains, whereas the other listeners' depths were smaller (Table II). This listener could also detect the small pulses at the largest modulation depth threshold for a 5-ms train in experiment 3. Also, listener CI1 showed very small modulation depths in both experiments. However, there is no similar relationship apparent for the normal-hearing listeners. The small number of subjects tested suggests that we should treat between-subject correlations with some caution. However, if it is indeed the case that performance in the two experiments correlates across implanted but not normal-hearing listeners, this may be due to the greater variability in peripheral processing across the former population.

Finally, we should note that detection thresholds did not vary significantly with the three different stimulus periods (4, 5, 6 ms) under test, for either group of listener (only for CI4). We can only speculate why the data of CI4 deviate from those of the other implantees. One reason may be that this subject has a different internal device (type 5.8, longitudinal) than the ones (5.6, radial); another reason may be that a fairly small range of periods was studied.

VI. DISCUSSION

A. Neural mechanisms

In the Introduction we described two mechanisms which could affect the perceived pitches of amplitude-modulated pulse trains in acoustic and in electric hearing. Our results are consistent with the operation of a mechanism, in which

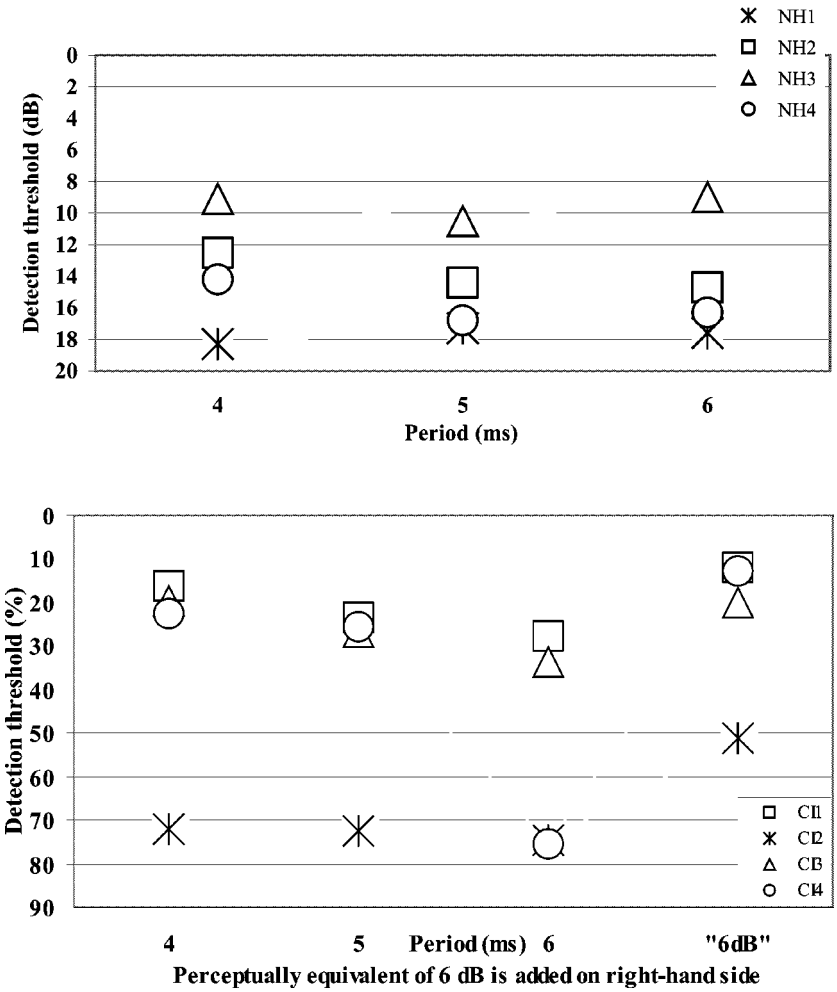


FIG. 4. Average detection thresholds for the smaller pulses at the three pulse rates, for the normal-hearing subjects (left) and the cochlear implantees (right). Thresholds indicate the value at which the modulated stimulus cannot be distinguished from the unmodulated one.

the high-amplitude pulses reduce the probability of auditory nerve (AN) firing in response to subsequent low-amplitude pulses, with this effect being stronger when the interval between the large and subsequent small pulses is 4 ms (condition A) than when it is 6 ms (condition B). This refractory effect, which can be thought of as a form of partial masking, would be expected to decrease when the interpulse intervals are increased overall (e.g., to 8 and 12 ms), causing the pitches of the stimuli in conditions A and B to become more similar. This pattern of results was indeed observed in experiment 2.

Because the delays between the high- and low-amplitude pulses differed only by 2 ms between conditions A and B of experiment 1, and because pitch differences were observed in electric as well as in acoustic hearing, we believe our findings can be explained in terms of short-term refractory effects. Although we cannot rule out an effect of refractoriness in neural structures central to the AN, a consideration of the time constants of AN refractoriness is generally consistent with our data. Dynes (1996) measured the threshold for a cat AN fiber in response to a single 100-μs monophasic cathodic pulse presented after a supra-threshold pulse, as a function of interpulse delay. He found that threshold dropped markedly back to resting threshold as interpulse delay was increased from 0 to 5 ms. Similarly, Fitzpatrick *et al.* (1999) observed that the probability of a cat AN fiber responding to a 30-μs acoustic click, preceded by another click of the same amplitude, increased markedly as the interpulse interval was increased from 1 to 8 ms, with a 15% increase occurring between intervals of 4 and 6 ms. Hence in both modalities, the AN will have shown greater recovery after an interpulse interval of 6 ms than after an interval of only 4 ms. Data of the electrically evoked whole nerve action potential (EAP) recorded from Ineraid CI-users show that the recovery of peripheral neural response is generally complete within 6 ms (Brown *et al.*, 1990, 1996). These experiments also demonstrate that the refractory properties of the auditory nerve vary across subjects and across electrodes. EAP data of cats are similar with full recovery occurring at an interpulse interval of 8 ms, while those of guinea pigs demonstrate recovery times in the order of 16 ms (Matsuoka *et al.*, 2000). In an earlier study, EAP data of Wilson *et al.* (1994), obtained with human implantees, demonstrated alternations in the response amplitude to successive pulses in a constant-amplitude pulse train. This effect, which was attributed to neurons being in a partial refractory state for every other pulse, occurred when the pulse rate exceeded about 400 pps (2.5 ms).

If we make the reasonable assumption that the effects of refractoriness on neural responses also depends on the relative amplitudes of successive pulses, then one can predict that the AN response to the low-amplitude pulses in condition A (Fig. 1) will be less likely to cause a neural firing than

the low-amplitude pulses in condition B. Furthermore, because AN recovery will have reached an asymptote by the longer delays used in experiment 2 (Dynes and Delgutte, 1992; Fitzpatrick *et al.*, 1999), this explanation is consistent with the absence of an A-B pitch difference at those delays.

Finally, it is worth considering our explanation for the differences in pitch between the A and B modulated stimuli, in the light of Carlyon *et al.*'s (2002) study of the pitch of unmodulated pulse trains. They presented subjects with a train of equal-amplitude pulses in which the interpulse interval (IPI) alternated between 4 and 6 ms, and found that its pitch equalled that of an isochronous pulse train with an IPI of 5.7 ms. They modeled this and other findings using a weighted sum of first-order IPIs, in which the 6-ms IPI received a much higher weight than the 4-ms IPI. They also concluded that this selective weighting of longer IPIs was too great to be *solely* accounted for by recovery functions of AN fibers (Fitzpatrick *et al.*, 1999; Kiang, 1965). The present results are consistent with this interpretation, but suggest that refractoriness can have at least some effect on temporal pitch perception, especially when the pulses have unequal amplitudes. One way in which this could occur would be by refractoriness AN fibers to ''miss'' the occasional low-amplitude pulse, leading to a (neural) interspike interval of $10\,(=4+6)$ ms.

B. Summary

The experiments described here measured the effect of temporal manipulations on pitch perception in both acoustic and electric hearing. The results extend the findings of Carlyon *et al.* (2002), who measured the pitch of equal-amplitude pulse trains in which the interpulse intervals alternated between 4 and 6 ms, to include amplitude-modulated pulse trains, in which the amplitude of every other pulse was attenuated. The results showed that when the low-amplitude pulses occurred at the end of the 4-ms interpulse intervals (''condition A''), the pitch was lower than when they occurred at the end of the 6-ms intervals (''condition B,'' see Fig. 1). The sign of this pitch difference, combined with the fact that it disappeared when interpulse intervals were doubled or trebled, is consistent with an effect of AN refractoriness on pitch perception. If this interpretation is correct, then our findings provide a behavioral, supra-threshold measure of refractoriness, which can be applied to both normal-hearing and clinical populations.

The pitch difference between conditions A and B, together with the overall drop in pitch observed with increasing modulation depth, was similar for both normal-hearing and implanted listeners. This adds to existing evidence that bandpass-filtered acoustic pulse trains can, under appropriate conditions, provide a good analog of electric hearing (McKay and Carlyon, 1999; Carlyon *et al.*, 2002). This combined approach has helped the development of an account of the factors that do and do not affect the pitch of electric and acoustic pulse trains. Specifically pitch is strongly affected by the longer first-order intervals present in the stimulus (Carlyon *et al.*, 2002), by modulation depth (McKay *et al.*, 1995; McKay and McDermott, 1999), and by refractoriness of lower-amplitude pulses by larger pulses (present study). There is no evidence for an effect of higher-order intervals (such as between one pulse and the next-but-one), and very short first-order intervals have only a small influence (Carlyon *et al.*, 2002). Our results are consistent with a mechanism that analyzes first-order intervals, if we assume that this analysis is performed separately on different groups of neurons, where the neurons within each group have a similar CF. Those groups of neurons that are remote from the stimulating electrode will respond only to the higher-amplitude pulses, whereas groups of neurons closer to the electrode will respond to both small and big pulses. The pitch mechanism could then weight the different groups; because the proportion of neurons responding to the smaller pulses will decrease with increasing modulation depth or increasing refractoriness, this would lead to a reduction in estimated pitch. If correct, this should help us to predict the pitch of arbitrary patterns of stimulation, such as might emerge from the output of a cochlear implant speech processor in response to a mixture of voices. In speech-processing strategies such as CIS and SPEAK, such patterns would be applied to the *envelopes* imposed on fixed-rate pulse trains on *multiple* channels, rather than being reflected in the intervals between *individual* pulses on a *single* channel, as studied here. Nevertheless, it is likely that a detailed account based on our relatively simple stimuli can be extended to encompass these more complex ''real-life'' situations.

ACKNOWLEDGMENTS

We sincerely thank the normal-hearing and implanted subjects who spent many hours picking the higher of two ''bleeps.'' The acoustical experiments were performed by members of the Lab. Exp. ORL (Leuven, Belgium); the electrical experiments were performed by subjects implanted at the University ENT-Department of the A.Z. St. Augustinus Hospital in Antwerp, or at the University Hospital in Leuven. We also thank Dr. Ian Nimmo-Smith (MRC, Cambridge) for help with the statistical analyses. This study was supported by the Research Fund of the Katholieke Universiteit Leuven and by the Fund for Scientific Research—Flanders (Belgium).

Brown, C. J., Abbas, P. J., and Gantz, B. (**1990**). ''Electrically evoked whole-nerve action potentials: data from human cochlear implant users,'' J. Acoust. Soc. Am. **88**, 1385–1391.

Brown, C. J., Abbas, P. J., Borland, J., and Bertschy, M. R. (**1996**). ''Electrically evoked whole nerve action potentials in Ineraid cochlear implant user: responses to different stimulating electrode configurations and comparison to psychophysical responses,'' J. Speech Hear. Res. **39**, 453–467.

Carlyon, R. P. (**1996**). ''Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker,'' J. Acoust. Soc. Am. **99**, 517–524.

Carlyon, R. P. (**1997**). ''The effects of two temporal cues on pitch judgments,'' J. Acoust. Soc. Am. **102**, 1097–1105.

Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). ''Temporal pitch mechanisms in acoustic and electric hearing,'' J. Acoust. Soc. Am. **112**, 621–633.

Dynes, S. B. C., and Delgutte, B. (**1992**). ''Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea,'' Hear. Res. **58**, 79–90.

Dynes, S. B. C. (**1996**). ''Discharge Characteristics of Auditory Nerve Fibers for Pulsatile Electrical Stimuli,'' Ph.D. thesis, Massachusetts Institute of Technology.

Fitzpatrick, D. C., Kuwada, S., Kim, D. O., Parham, K., and Batra, R. (**1999**). ''Responses of neurons to click-pairs as simulated echoes: Auditory nerve to auditory cortex,'' J. Acoust. Soc. Am. **106**, 3460–3472.

Geurts, L., and Wouters, J. (**2000**). ''A concept for a research tool for experiments with cochlear implant users,'' J. Acoust. Soc. Am. **108**, 2949–2956.

Javel, E., and Shepherd, R. K. (**2000**). ''Electrical stimulation of the auditory nerve. III. Response initiation sites and temporal fine structure,'' Hear. Res. **140**, 45–76.

Kiang, N. Y. S. (**1965**). Discharge Patterns of Single Fibers in the Cat's Auditory Nerve (MIT, Cambridge, MA).

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). ''The lower limit of pitch as determined by rate discrimination,'' J. Acoust. Soc. Am. **108**, 1170–1180.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Matsuoka, A. J., Abbas, P. J., Rubinstein, J. T., and Miller, C. A. (**2000**). ''The neuronal response to electrical constant-amplitude pulse train stimulation: evoked compound action potential recordings,'' Hear. Res. **149**, 115–128.

McDermott, H. J., and McKay, C. M. (**1997**). ''Musical pitch perception with electrical stimulation of the cochlea,'' J. Acoust. Soc. Am. **101**, 1622–1631.

McKay, C. M., and Carlyon, R. P. (**1999**). ''Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains,'' J. Acoust. Soc. Am. **105**, 347–357.

McKay, C. M., and McDermott, H. J. (**1996**). ''The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites,'' J. Acoust. Soc. Am. **100**, 1081–1092.

McKay, C. M., and McDermott, H. J. (**1998**). ''Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals,'' J. Acoust. Soc. Am. **104**, 1061–1074.

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

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1995**). ''Pitch matching of amplitude modulated current pulse trains in cochlear implantees: the effect of modulation depth,'' J. Acoust. Soc. Am. **97**, 1777–1785.

Meddis, R., and Hewitt, M. (**1991**). ''Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery. I: pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.

Peeters, S., Offeciers, F. E., Joris, P., and Moeneclaey, L. (**1993**). ''The Laura cochlear implant programmed with the continuous interleaved and phase-locked continuous interleaved strategies,'' in Cochlear Implants: New Perspectives, edited by B. Fraysse and O. Deguine, Adv. Otorhinolaryngol. **48**, 261–268 .

Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.

SPSS 10.0 (**2000**). SPSS, Inc., Chicago, IL.

van den Honert, C., and Stypulkowski, P. H. (**1984**). ''Physiological properties of the electrically stimulated auditory nerve. II. Single fiber recordings,'' Hear. Res. **14**, 225–243.

van Wieringen, A., and Wouters, J. (**1999**). ''Natural vowel and consonant recognition by Laura cochlear implantees,'' Ear Hear. **20**, 89–103.

Wilson, B. S., Finley, C. C., Zerbi, M., and Lawson, D. T. (**1994**). ''Speech processors for auditory prostheses,'' Seventh Quartely Progress Report, NIH Contract No. N01-DC-2-2401, Center for auditory prosthesis research.

Yoshie, N. (**1968**). ''Auditory nerve action potential responses to clicks in man,'' Laryngoscope **78**, 198–213.

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

Extending the domain of center frequencies for the compressive gammachirp auditory filter

Roy D. Patterson[a)]
Centre for Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

Masashi Unoki[b)]
Centre for Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom and Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Tatsunokuchi-machi, Nomi-gun, Ishikawa 923-1292, Japan

Toshio Irino[c)]
Faculty of Systems Engineering, Wakayama University/NTT Communication Science Laboratories, NTT Corp/CREST-JST, 930 Sakaedani, Wakayama 640-8510, Japan

(Received 19 November 2001; revised 8 May 2003; accepted 17 June 2003)

The gamma*tone* filter was imported from auditory physiology to provide a time-domain version of the roex auditory filter and enable the development of a realistic auditory filterbank for models of auditory perception [Patterson *et al.*, J. Acoust. Soc. Am. **98**, 1890–1894 (1995)]. The gamma*chirp* auditory filter was developed to extend the domain of the gammatone auditory filter and simulate the changes in filter shape that occur with changes in stimulus level. Initially, the gammachirp filter was limited to center frequencies in the 2.0-kHz region where there were sufficient ''notched-noise'' masking data to define its parameters accurately. Recently, however, the range of the masking data has been extended in two massive studies. This paper reports how a compressive version of the gammachirp auditory filter was fitted to these new data sets to define the filter parameters over the extended frequency range. The results show that the shape of the filter can be specified for the entire domain of the data using just six constants (center frequencies from 0.25 to 6.0 kHz and levels from 30 to 80 dB SPL). The compressive, gammachirp auditory filter also has the advantage of being consistent with physiological studies of cochlear filtering insofar as the compression of the filter is mainly limited to the passband and the form of the chirp in the impulse response is largely *in*dependent of level. [DOI: 10.1121/1.1600720]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.64.Bt [MRL]

I. INTRODUCTION

Baker *et al.* (1998) and Glasberg and Moore (2000) have recently measured threshold for a sinusoid in the presence of an asymmetric notched noise (Patterson and Nimmo-Smith, 1980) over what is effectively the entire range of frequencies and levels encountered in everyday hearing, and used the roex filter function (Patterson *et al.*, 1982) to estimate the shape of the human auditory filter over this comprehensive range of frequencies and levels. They also took particular care with details such as the headphone response and the effect of the outer and middle ears. The studies both showed that (a) the shape of the auditory filter changes with level at all center frequencies; (b) the gain at the peak frequency increases as stimulus level *decreases*; and (c) the degree of compression at the center frequency increases with signal frequency in the range 0.25 to 1.0 kHz.

The roex auditory filter provided excellent fits to the data, but the phase response of the roex auditory filter is not defined. This means that there is no impulse response for the roex auditory filter and no time-domain filterbank for auditory modeling or speech preprocessing. This was the original motivation for the development of the gammatone auditory filterbank described in Patterson *et al.* (1995). In this paper, we fit the compressive gammachirp auditory filter of Irino and Patterson (2001) to the human masking data of Baker *et al.* (1998) and Glasberg and Moore (2000) to demonstrate that it can explain the data and so provide the basis for a time-domain, level-dependent filterbank for auditory modeling and speech preprocessing. There are now a number of descendants of the original gammatone filterbank, including some very recent ones (e.g., Meddis *et al.*, 2001; Zhang *et al.*, 2001) developed to characterize cochlear physiology.[1] A brief summary of the development of the gammatone and gammachirp filterbanks over the past 30 years is provided in Appendix A.

II. THE COMPRESSIVE GAMMACHIRP AUDITORY FILTER

In general, the auditory filter has been derived using data from the asymmetric-notched-noise experiment (Patterson and Nimmo-Smith, 1980), the power spectrum model of masking (Fletcher, 1940; Patterson, 1976), and a flexible representation of the filter shape with the basic assumptions about the architecture of the filter. This section describes the

a)Electronic mail: roy.patterson@mrc-cbu.cam.ac.uk
b)Electronic mail: unoki@jaist.ac.jp
c)Electronic mail: irino@sys.wakayama-u.ac.jp

latest methods for deriving an auditory filter from notched-noise data and how they are applied in the case of the gammachirp auditory filter.

A. The notched-noise experiment

In the notched-noise experiment, the listener is required to detect a brief sinusoidal signal, referred to as a ''probe'' tone, in the presence of a masker which is a noise with a spectral notch in the frequency region of the probe tone. This ''notched noise'' has a constant spectrum level N_0, in a band below the tone between $f_{l_{\min}}$ and $f_{l_{\max}}$ and in a band above the tone between $f_{u_{\min}}$ and $f_{u_{\max}}$. The probe tone is varied to determine the power required to make the probe just audible (probe ''threshold''), as a function of the width of the notch in the noise. It is assumed that the observer listens for the tone through an auditory filter centered in the region of the tone to maximize the ratio of signal power to noise power at the output of the filter. In the asymmetric, notched-noise experiment, for each combination of signal frequency and notch width, the center of the notch is varied relative to the signal frequency, and the threshold differences observed in conditions with mirror-image asymmetry are used to define the asymmetry of the passband and/or tails of the auditory filter (Patterson and Nimmo-Smith, 1980). The dynamic range of the threshold measurements can be up to 60 dB, and this means that the procedure is highly sensitive to shape even in the tails of the filter.[1] The dynamic range of Fletcher's (1940) band-widening experiment is limited to about 5 dB, and so it provides little information about filter shape both in the tails and in the passband (Patterson and Henning, 1977).

B. The power spectrum model of masking

If the slope of the noise cutoff is very steep, it is possible to write a function that relates probe threshold to the integral of the auditory filter. This relationship provides the basis for the fitting procedure used to derive the filter shape (Patterson, 1974). If the ''shape'' of the auditory filter (that is, its magnitude spectrum) is represented by the weighting function, $W(f)$, then the threshold predicted by the power spectrum model of masking is given by

$$P_s = K \cdot \left(\int_{f_{l_{\min}}}^{f_{l_{\max}}} N_0 \cdot W(f)\,df + \int_{f_{u_{\min}}}^{f_{u_{\max}}} N_0 \cdot W(f)\,df \right), \quad (1)$$

where P_s is the power of the probe tone at threshold, N_0 is the spectrum level of the noise, and K is a constant which is related to the efficiency of the detection mechanism following the auditory filter. In the limits on the filter integrals (f_l and f_u), the subscripts l and u refer to the *l*ower and *u*pper noise bands; the subscripts min and max refer to the cutoff frequencies of the noise bands. It is referred to as a ''power-spectrum'' model because it includes the simplifying assumption that the fluctuations in the noise can be ignored. More comprehensive models of masking that include either the energy statistics or envelope fluctuations are described in Patterson and Henning (1977) and Derleth and Dau (2000), respectively. However, in the asymmetric notched-noise experiment, it is essentially the filter shape itself that determines threshold, and so the current paper is limited to the power-spectrum model. Once a candidate filter shape is defined mathematically, it can be used in conjunction with the threshold data and Eq. (1) to determine the parameters of the filter.

C. The compressive gammachirp filter function

The complex form of the gammachirp auditory filter (Irino and Patterson, 1997) is

$$g_c(t) = a t^{n_1 - 1} \exp(-2\pi b_1 \,\mathrm{ERB}(f_{r1}) t) \times \exp(j 2\pi f_{r1} t + j c_1 \ln t + j\phi_1), \quad (2)$$

where time $t > 0$; a is amplitude; n_1 and b_1 are parameters defining the envelope of the gamma distribution; c_1 is the chirp factor; f_{r1} is the asymptotic center frequency; $\mathrm{ERB}(f_{r1})$ is the equivalent rectangular bandwidth (Glasberg and Moore, 1990); ϕ_1 is the initial phase; and $\ln t$ is the natural logarithm of time. When $c_1 = 0$, Eq. (2) reduces to the complex impulse response of the gammatone filter (Patterson *et al.*, 1995). The Fourier magnitude spectrum of the gamma*chirp* filter is

$$|G_C(f)| = a_\Gamma \cdot |G_T(f)| \cdot \exp(c_1 \theta_1(f)), \quad (3)$$

$$\theta_1(f) = \arctan\left(\frac{f - f_{r1}}{b_1 \,\mathrm{ERB}(f_{r1})} \right). \quad (4)$$

$|G_T(f)|$ is the Fourier magnitude spectrum of the gamma*tone* filter, and $\exp(c_1\theta_1(f))$ is an asymmetric function since θ_1 is an antisymmetric function centered at the asymptotic frequency, f_{r1} [Eq. (4)]. a_Γ is the relative amplitude of the magnitude spectrum of the gamma*tone* filter. This is the ''analytic'' form of the gammachirp filter in which the chirp varies with level (see Appendix A of Irino and Patterson, 1997).

Irino and Patterson (2001) decomposed the asymmetric function, $\exp(c_1\theta_1(f))$, into separate low-pass and high-pass asymmetric functions in order to represent the passive basilar membrane and the subsequent level-dependent component separately in the filter function. The resulting ''compressive'' gammachirp filter, $|G_{CC}(f)|$, is

$$|G_{CC}(f)| = [a_\Gamma |G_T(f)| \cdot \exp(c_1\theta_1(f))] \cdot \exp(c_2\theta_2(f))$$
$$|G_{CC}(f)| = |G_{CP}(f)| \cdot \exp(c_2\theta_2(f)). \quad (5)$$

Conceptually, this compressive gammachirp is composed of a level-*in*dependent, ''passive'' gammachirp filter, $|G_{CP}(f)|$, that represents the passive basilar membrane, and a level-dependent, high-pass asymmetric function, $\exp(c_2\theta_2(f))$, that represents the active mechanism in the cochlea. The architecture of the compressive gammachirp is illustrated in Fig. 1. The upper panel shows how a passive gammachirp (GC, solid line) is constructed by cascading a gamma*tone* filter (GT, dashed and dotted line) with a low-pass asymmetric function (LP-AF, dashed line). Since the gain scales are logarithmic, adding the values of the component functions, GT and LP-AF, produces the value of the composite function, GC. This passive GC filter largely determines the form of the

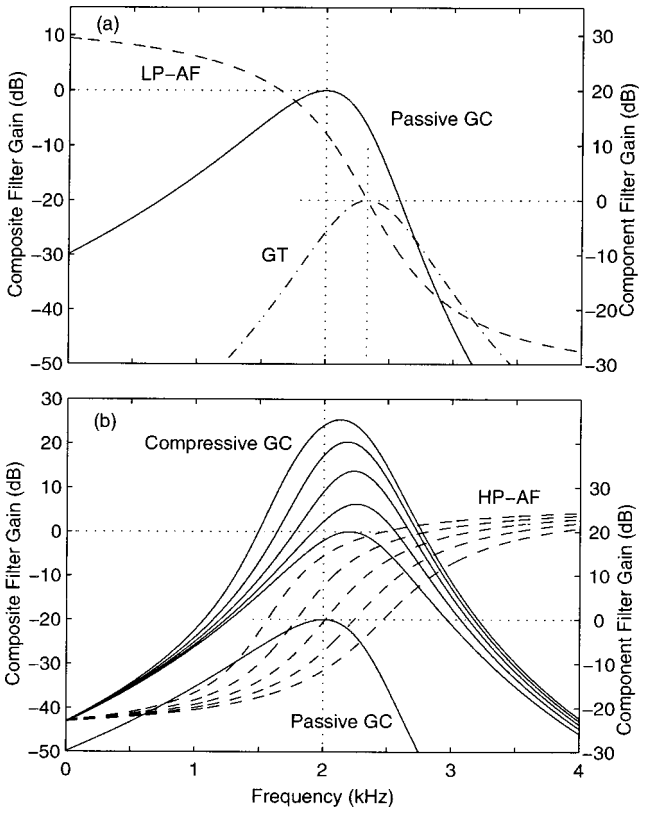


FIG. 1. Illustration of (a) the construction of an asymmetric, passive gammachirp filter (GC) from a symmetric gammatone filter (GT) and a low-pass asymmetric function (LP-AF), and (b) the construction of a set of compressive gammachirp filters (GC) from one passive GC filter and a high-pass asymmetric function (HP-AF) whose center frequency shifts up as stimulus level increases. The vertical dotted lines show the peak frequencies of (a) the passive GC and GT filters (left and right, respectively) and (b) the passive GC filter.

chirp, and so, the chirp no longer varies significantly with level.

The lower panel illustrates how a level-dependent set of compressive GC filters (upper set of solid lines; left ordinate) can be produced by cascading a fixed passive GC filter (lower solid line; right ordinate) with a set of high-pass asymmetric functions (HP-AF) (set of dashed lines; right ordinate). When the leftmost HP-AF is cascaded with the passive GC, it produces a compressive GC filter with much more gain, than when the rightmost HP-AF is cascaded with the passive GC. As a result, at the peak frequency of the compressive GC, gain *decreases* as level increases. The ordinate is filter gain and, following the physiological convention, the gain is normalized to the peak value of the filter associated with the highest probe level, which in this case is 70 dB. The five filter shapes have been calculated for probe levels of 30, 40, 50, 60, and 70 dB. The filter is referred to as a ''compressive'' gammachirp because the compression around the peak frequency is incorporated into the filtering process itself. The high-pass asymmetric function makes the passband of the composite gammachirp more symmetric at lower levels.

Mathematically, the angular variables are rewritten in terms of the center frequency and bandwidth of the passive gammachirp filter and the level-dependent asymmetric function to accommodate the shifting of the asymmetric function relative to the basilar-membrane function with level. If the filter center frequencies are f_{r1} and f_{r2}, respectively, then from Eq. (4)

$$\theta_1(f) = \arctan\left(\frac{f - f_{r1}}{b_1\,\mathrm{ERB}(f_{r1})}\right)$$

and

$$\theta_2(f) = \arctan\left(\frac{f - f_{r2}}{b_2\,\mathrm{ERB}(f_{r2})}\right). \tag{6}$$

The peak frequency of the passive gammachirp, $|G_{CP}(f)|$, is

$$f_{p1} = f_{r1} + c_1 b_1\,\mathrm{ERB}(f_{r1})/n_1. \tag{7}$$

In this form, the chirp parameters, c_1 and c_2, can be fixed with, for example, $c_1 = -2.5$ and $c_2 = 2.0$ as shown in Fig. 1, and the level dependency can be associated with the center frequency of the *high-pass* asymmetric function, f_{r2}. The center frequency of the low-pass asymmetric function, f_{r1}, is fixed, so that the peak frequency of the gammachirp that represents the passive basilar membrane is fixed.

D. Refinement of the power spectrum model of masking

Conceptually, the terms in the equation for the power spectrum model of masking [Eq. (1)] refer to values at the output of the auditory filter. Originally, the values of P_s and N_0 at the input to the ear canal were used as estimates of the internal values, and the only aspect of auditory function that was considered was the transfer function applied by the auditory filter to the two noise bands. Much of the subsequent progress in the measurement of auditory filter shape can be regarded as progress in specifying how the masker bands *and* the signal are transformed by the outer and middle ear and by the auditory filter, as the stimuli progress from the air to the *output* of the auditory filter where threshold is actually determined. This section reviews the refinements and how they are implemented in the current study.

The outer and middle ear both affect the spectrum of the masker and thus the levels of the noise bands at the output of the auditory filter, especially when the probe frequency is either low or high. Rosen and Baker (1994) and Rosen *et al.* (1998) demonstrated that including a transfer function that represented precochlear filtering reduced the rms error of fits in multilevel filter-shape studies. More recently, Glasberg and Moore (2000) showed how the effects of precochlear filtering can be neutralized by tailoring the masker to produce uniformly exciting noise (UEN) at the input to the cochlea. In the current paper, when fitting the data of Baker *et al.* (1998) and Glasberg and Moore (2000), the effects of precochlear processing are included in the same way as they were in the original studies.

With regard to signal level, Patterson and Nimmo-Smith (1980) pointed out that, in asymmetric notched noises, when the auditory filter shifts off-probe frequency to improve the signal-to-noise ratio, the filter attenuates the signal as well as the noise bands. The effects of off-frequency listening are

included in the current study by locating the auditory filter that produces the best signal-to-noise ratio when calculating threshold estimates.

The studies of Rosen *et al.* (1998) and Glasberg and Moore (2000) showed that it is also important to include a low-level limit on threshold since the experiments include conditions where threshold approaches absolute threshold. This term is also included in the current fits; it is designated P_0 and it is an additive constant in linear power terms.

The compression applied in the filtering process has a major effect on stimulus level at the output of the filter. In the compressive GC filter, the last linear element is the passive GC and, in the fitting process, it is the total level of the probe plus the two masker bands at the output of passive GC that controls the position of the HP-AF, and thus the compression and gain of the composite gammachirp filter. The level at the output of the passive GC is designated P_{gcp}.

Despite the refinement of the power-spectrum model, the basic technique for deriving the auditory filter shape remains the same; probe level at threshold is used in conjunction with carefully designed noise maskers to estimate the integral of the auditory filter in a set of relevant conditions. Then, a function like the gammachirp is selected to represent the auditory filter and the threshold data are used to estimate the best values for the filter parameters. For further details see Patterson and Moore (1986) and Irino and Patterson (1997).

Irino and Patterson (2001) demonstrated how the compressive gammachirp filter could be fitted to large sets of level-dependent physiological and psychophysical data when they are all associated with one probe frequency. In the next section, the technique is applied to the data associated with all the individual probe frequencies in Baker *et al.* (1998) and Glasberg and Moore (2000). Then, the technique is extended, first to fit the data from multiple probe frequencies simultaneously (within a single study), and then to produce one global fit to all of the data in both studies.

III. FITTING TO INDIVIDUAL PROBE FREQUENCIES

Initially, the gammachirp filter was fitted to all of the data associated with the individual probe frequenies in each study. Seven parameters were used to fit the data: five gammachirp filter parameters, b_1, c_1, b_2, c_2, and f_{rat} [f_{rat} is f_{r2}/f_{p1}, see Eq. (9) in Irino and Patterson, 2001]; and two nonfilter parameters K and P_0. The order of the gamma function, n_1, was fixed at 4 as in previous studies (de Boer and de Jongh, 1978; Carney and Yin, 1988; Irino and Patterson, 1997, 2001). Initial studies indicated that a good fit could be obtained for the data sets associated with all of the individual probe frequencies in both studies provided one parameter, f_{rat}, was allowed to vary linearly with level. It controls the frequency shift of the HP-AF relative to the passive GC and it is determined by P_{gcp}, the level of the signal plus masker at the output of the passive GC filter. Specifically

$$f_{rat} = f_{rat}^{(0)} + f_{rat}^{(1)} \cdot P_{gcp}. \tag{8}$$

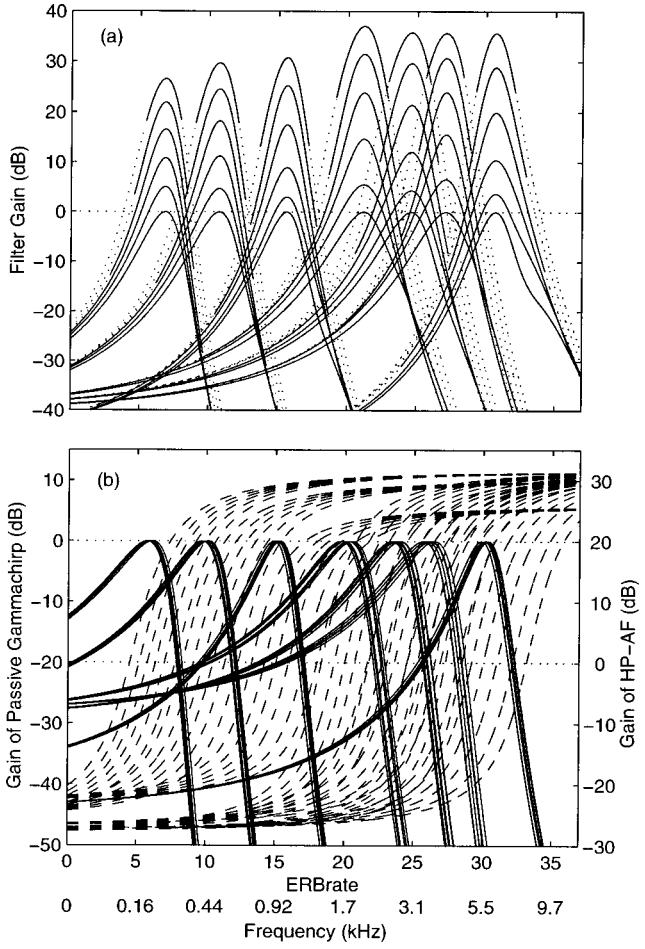


FIG. 2. Families of compressive gammachirp filters (a) and their component filters (b) for the threshold data of Baker *et al.* (1998); the seven probe frequencies are 0.25, 0.5, 1.0, 2.0 3.0, 4.0, and 6.0 kHz. The six curves within each set show how the function changes as probe level increases from 30 to 80 dB in 10-dB steps. The thresholds associated with each probe frequency were fitted individually, and the complete set of fits involved 42 filter coefficients.

The superscripts 0 and 1 designate the intercept and slope of the line. This produces an eight-coefficient model with one coefficient for each parameter except f_{rat}, which has two coefficients. A detailed description of the steps in the fitting procedure is presented in Appendix B.

A. The data of Baker *et al.* (1998)

1. The composite filters

When the eight-coefficient model is fitted separately to the data for the seven probe frequencies of Baker *et al.* (1998), the result is the seven families of compressive GC filters shown in Fig. 2(a). The six functions in each family show filters at probe levels from 30 to 80 dB in 10-dB steps. The families of filters all have the general form illustrated in Fig. 1(b); the gain of the composite GC *decreases* monotonically with increasing stimulus level in the passband of the filter, as would be expected. The range of the gain, or compression, is a little under 30 dB at the lower probe frequencies (0.25, 0.5, and 1.0 kHz) and a little over 30 dB at the higher probe frequencies (2.0, 3.0, 4.0, and 6.0 kHz). The

range of gain is, in each case, a little greater than that reported by Baker *et al.* (1998) for the same data. The bandwidth increases slowly with level.

There are three aspects of the figure that should be noted: (1) The abscissa is ERBrate, that is, signal frequency measured in terms of filter bandwidths (Glasberg and Moore, 1990). It is a ''critical band'' scale that provides a more uniform representation of cochlear place, or hair cell density, than a log-frequency scale; in humans, one ERB occupies approximately 0.9 mm of the cochlear partition (Moore, 1997, p. 115). On this scale, the bandwidth of the gammachirp filter is roughly constant across probe frequency. The corresponding frequency values are presented below the ERBrate values. (2) Each set of filters has the same physiological format as described in the introduction to the compressive GC filter (Sec. II C). Normalizing the gain of the filters to the peak value of the filter associated with the *highest* probe level results in a figure which gives the impression that we can specify the form of the *lower-level* filters over a range of as much as 80 dB. This is clearly not the case; there are no threshold values from the notch-noise experiment to specify the tails of the lower level filters. Accordingly, the line used to represent each composite filter changes its form from solid to dotted at the point where the filter function would intersect the threshold limit, P_0. (3) The center frequency of the composite filters in Fig. 2(a) does not vary with level within a given set. This contrasts with figures from physiological studies which show the peak of the physiological response shifting downwards in frequency as level increases (e.g., Ruggero, 1992). In fact, the two representations are different and there is no fundamental discrepancy. The important point to note is that the psychophysical paradigm forces the listener to use the cochlear filter that produces the best signal-to-noise ratio at all times, and as a result, it is not possible to extract information about the change in position of one particular cochlear filter with level from these data. The data do contain information about the relative position of the HP-AF with respect to the passive gammachirp, and this is why we use the variable f_{rat} rather than f_{r2} and f_{p1}. Each of the families of composite gammachirp filters in Fig. 2(a) shows the filters associated with the probe frequency as level increases, rather than the response of one cochlear filter to an increase in level. This is also the reason we refer to these sets of filters in terms of their ''probe frequency'' rather than referring to their center frequency. The reader should note, however, that in most masking experiments a given probe frequency may be associated with a range of filter peak frequencies, and vice versa.

2. The component filters and their coefficients

The passive GC filters and HP-AFs that are cascaded to produce the composite filters in Fig. 2(a) are presented in Fig. 2(b) by solid and dashed lines, respectively. The abscissa is ERBrate as in Fig. 2(a); the ordinate is component filter gain with the values for the passive GCs on the left and those for the HP-AFs on the right. The level dependency appears in the frequency shift of the passive GC relative to the corresponding HP-AF, within each family. On the ERBrate scale, the passive GCs and the HP-AFs have roughly the same shape across the full range of probe frequencies. The passive GCs are a little more asymmetric at 2.0, 3.0, and 4.0 kHz than at the other probe frequencies, and the range of the HP-AFs is somewhat less. However, these factors counteract each other when the component filters are cascaded, and so the shape of the composite GC does not vary much with probe frequency.[2]

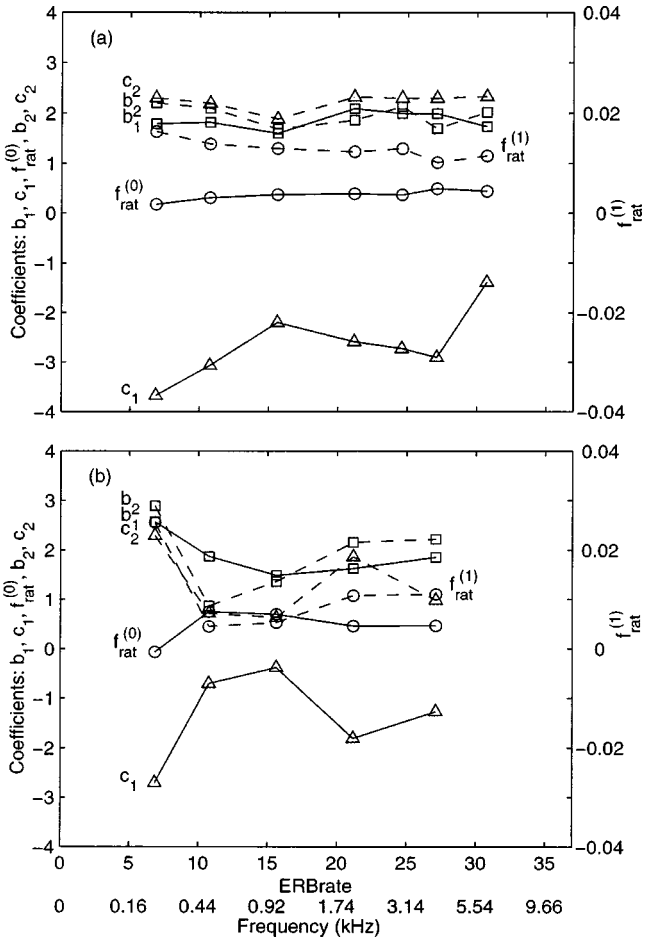


FIG. 3. Filter coefficients for the individual fits to the data of (a) Baker *et al.* (1998) and (b) Glasberg and Moore (2000), plotted as a function of probe frequency on an ERBrate scale.

The filter coefficients are plotted as a function of probe frequency in Fig. 3(a). Five of the six functions are smooth and essentially flat, indicating that the variation in coefficient value with frequency is largely specified by ERBrate. The exception is c_1, which will be discussed shortly, following the description of the fit to the second data set.

B. The data of Glasberg and Moore (2000)

When the eight-coefficient model is fitted to the data for the five probe frequencies of Glasberg and Moore (2000), the result is the five families of compressive GC filters shown in Fig. 4(a); the format is the same as in Fig. 2(a). The families of filters have the same general form, and within each family, the gain of the composite gammachirp increases monotonically with decreasing stimulus level in the passband. For probe frequencies of 0.25 and 2.0 kHz, the tails of the compressive GCs are asymmetric and the gain range exceeds 20

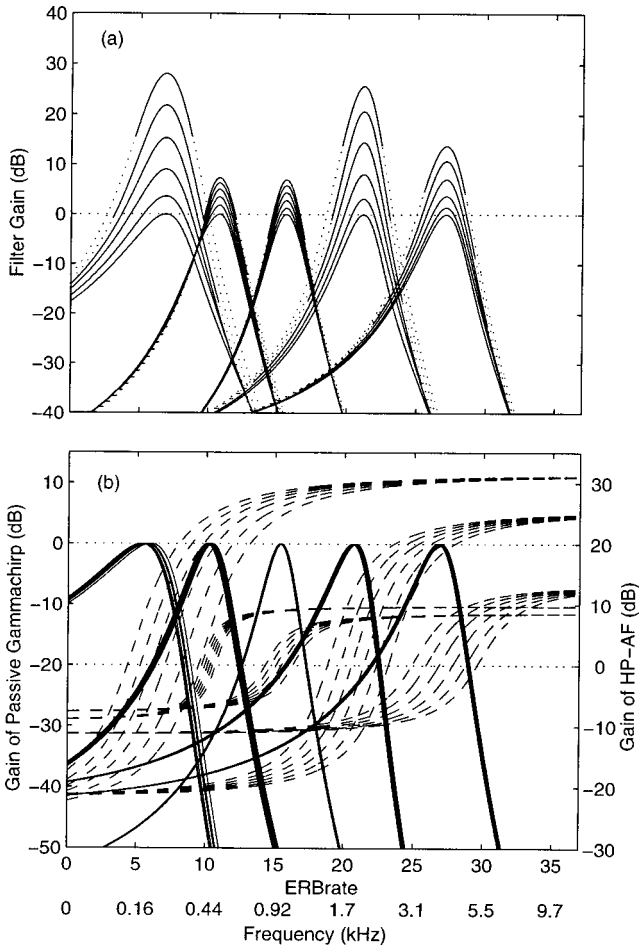


FIG. 4. Families of compressive gammachirp filters (a) and their component filters (b) for the threshold data of Glasberg and Moore (2000); the five probe frequencies were 0.25, 0.5, 1.0, 2.0, and 4.0 kHz. The six curves within each set of compressive GC filters (a) and HP-AFs (b) show how the functions change as probe level increases from 30 to 80 dB in 10-dB steps. The thresholds associated with each probe frequency were fitted individually, and the complete set of fits involved 30 filter coefficients.

dB. At 0.5, 1.0 and 4.0 kHz, however, the filters are largely *symmetric* and the gain range is too small to represent cochlear compression.

The passive GC filters and the HP-AFs for the data of Glasberg and Moore (2000) are presented in Fig. 4(b) by solid and dashed lines, respectively; the format is the same as in Fig. 2(b). At 0.25 and 2.0 kHz, the passive GCs are reasonably asymmetric and the HP-AFs have a substantial range. At 0.5 and 1.0 kHz, however, the passive GCs are too symmetric to represent the passive basilar membrane and the range of the HP-AFs is too small to represent cochlear compression.

The filter coefficients are plotted as a function of probe frequency in Fig. 3(b); the figure shows that the coefficients are more variable for this data set. Nevertheless, for the bandwidth and center-frequency coefficients (b_1, b_2, $f_{rat}^{(0)}$, and $f_{rat}^{(1)}$), the pattern of values is similar to that in Fig. 3(a) *and* the values are in the same absolute range. A comparison of the functions for the asymmetry coefficients in the data of Glasberg and Moore (2000), however, reveals a problem with fitting to individual probe frequencies. The variation in c_1 with probe frequency is negatively correlated with the variation in c_2. This suggests that there is a trade-off in the fitting process between the asymmetry of the passive GC and the asymmetry of the HP-AF with a relatively broad minimum. That is, an increase in the asymmetry of one can be counteracted by a decrease in the asymmetry of the other to produce a fit with effectively the same rms error. The fitting procedure locates the absolute minimum rms error at each probe frequency without regard to the values at neighboring probe frequencies, and in so doing often chooses combinations for adjacent probe frequencies that are highly unlikely. There is no reason to believe that the asymmetry of the passive basilar membrane changes rapidly with center frequency, and it is even less likely that the asymmetry of the active mechanism changes at the same point as the passive basilar membrane but in the opposite direction. It seems much more likely that the negative correlation between the asymmetry coefficients arises due to a lack of appropriate constraints on the fitting process.

IV. FITTING ALL PROBE FREQUENCIES FROM ONE STUDY SIMULTANEOUSLY

Baker *et al.* (1998) described a ''PolyFit'' procedure for fitting the roex auditory filter to data from multiple probe frequencies simultaneously. It was assumed that the variation in coefficient value across frequency could be represented by a regression line. Although the stated aim was to reduce the total number of free coefficients required to describe the data, the assumption will, at the same time, restrict the variation of coefficient value from one probe frequency to the next, and in so doing it can be expected to produce a more reasonable solution in terms of auditory physiology.

The PolyFit procedure of Baker *et al.* (1998) can be applied to the gammachirp auditory filter by representing each of the filter coefficients (b_1, c_1, b_2, c_2, $f_{rat}^{(0)}$, and $f_{rat}^{(1)}$) by a regression line with two regression coefficients (intercept and slope), and then varying these regression coefficients to determine the values that produce the minimum rms error for all of the thresholds at all of the probe frequencies and levels simultaneously. It is also necessary to provide regression functions for the nonfilter parameters, K and P_0, and here it is necessary to use a parabolic regression function because the values for the lower and higher probe frequencies are consistently greater than the values for the middle frequencies. Glasberg and Moore (2000) report a similar finding for these nonfilter parameters when fitting the roex filter to individual probe frequencies.

Three modifications were made to the PolyFit procedure as described by Baker *et al.* (1998); they are all comparatively minor but each improved the fit to some degree. First, the unit of the frequency scale for the regression lines and parabolas was switched from log frequency to ERBrate (as in Figs. 2–4). ERBrate is a better representation of the frequency/place map in the cochlea, and regression lines with this abscissa produced fits with smaller rms values. Specifically, we used a normalized version of ERBrate, E_f, that is

$$E_f = \mathrm{ERBrate}(f)/\mathrm{ERBrate}(1\ \ \mathrm{kHz}) - 1. \tag{9}$$

Although each filter coefficient was initially allowed to be level dependent, it soon became clear that only f_{rat} had to be written as a function of level to obtain a good fit, as observed previously in the fits to individual probe frequencies. The second modification was to use the Levenberg–Marquardt method (Press *et al.*, 1988) to minimize the rms difference between masked threshold (in dB) and the predicted value of masked threshold (also in dB). Finally, we noticed in preliminary simulations that the slope of the input/output function occasionally went slightly negative at higher stimulus levels; that is, the output actually decreased when the input increased. These unrealistic solutions arise when the fitting procedure uses passive gammachirp filters whose peak frequencies do not lie on the steep central section of the HP-AF. A smooth penalty function was introduced to avoid such fits, and the resulting procedure effectively matches the compression range of the HP-AF to that observed in the data. Glasberg *et al.* (1984) and Glasberg and Moore (2000) restricted the maximum compression allowed in their filter system for a similar reason.

A. The filters and regression lines for the simultaneous fits

The families of gammachirp filters for the simultaneous fit to the 973 thresholds of Baker *et al.* (1998) were similar to those shown in Fig. 2; the families at 2.0, 3.0, and 4.0 kHz became a little more symmetric while the others became a little less symmetric; the range of gain was about the same magnitude but the progression from family to family was smoother, as would be expected. The families of filters for the simultaneous fit to the 304 thresholds of Glasberg and Moore (2000) were all similar in shape to the 2.0-kHz family of individual filters in Fig. 4; the families at 0.25 and 4.0 kHz became somewhat more symmetric while those at 0.5 and 1.0 kHz became somewhat less symmetric; the range of gain was about 10 dB in all cases. Plots of the families of filters are omitted for brevity.

The regression lines for the filter coefficients from the simultaneous fits to the data of Baker *et al.* (1998) are presented in Fig. 5(a). They all have shallow slopes and they are essentially what one would predict from looking at the corresponding coefficient data in Fig. 3(a). The regression lines for the data of Glasberg and Moore (2000) are presented in Fig. 5(b). The lines for one bandwidth coefficient, b_1, and both center-frequency coefficients, $f_{rat}^{(0)}$ and $f_{rat}^{(1)}$, have shallow slopes and the functions are what would be expected from the data in Fig. 3(b). It is also the case that the slope and intercept values derived from the two studies are very similar for these three coefficients. The differences between the experiments are concentrated in the remaining coefficients: For the data of Glasberg and Moore, the regression lines for the second bandwidth coefficient, b_2, and the asymmetry coefficients, c_1 and c_2, have relatively shallow slopes [Fig. 5(b)], but the values are somewhat different from what might have been expected from the data in Fig. 3(b), and they are somewhat different from the values in Fig. 5(a). This means that the asymmetry trade-off is a little more complicated than described previously: In Fig. 3(b), not only is

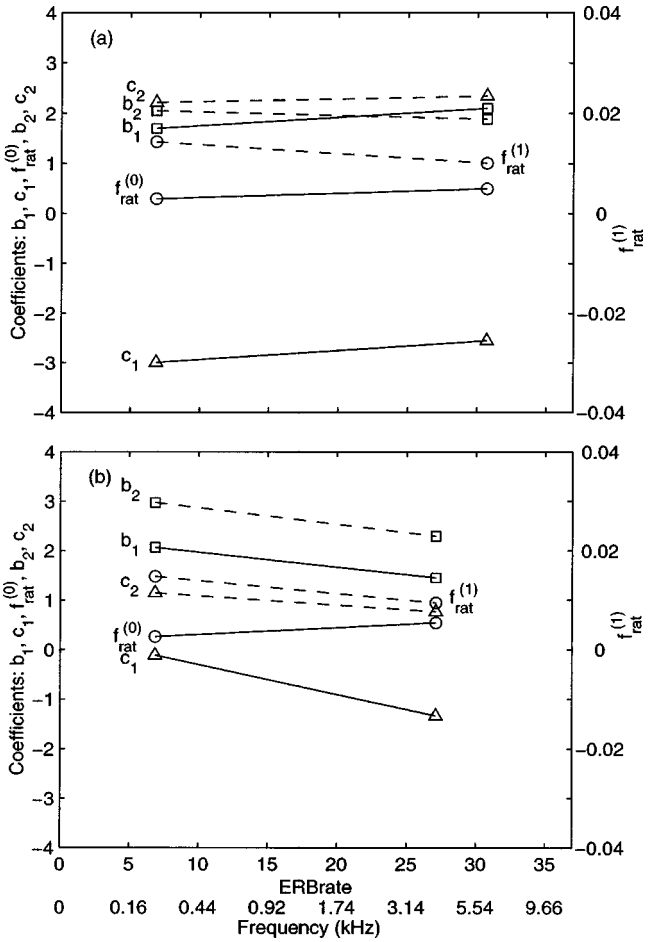


FIG. 5. Regression lines for the filter coefficients of the simultaneous fits to the data of (a) Baker *et al.* (1998) and (b) Glasberg and Moore (2000), plotted as a function of probe frequency on an ERBrate scale. The symbols on the regression lines are not data points; they are simply there to indicate which filter coefficient is represented.

c_1 negatively correlated with c_2 (triangles with solid and dashed lines); it is also negatively correlated with b_2 (squares with dashed line). Furthermore, b_2 is positively correlated with c_2. So, both the asymmetry and the position of the HP-AF adjust to changes in the asymmetry of the passive GC. In the simultaneous fit to the data of Glasberg and Moore, the result of this mutual adjustment is that the regression lines for b_2 and c_1 are higher than might have been expected, while that for c_2 is lower than might have been expected. Fortunately, there are no fundamental conflicts with the data of Baker *et al.* (1998). In Sec. V, we return to the topic of trade-offs and show that when the two data sets are combined, they lead to a surprisingly simple global fit.

B. The number of coefficients and the rms error

The individual fits to the data of Baker *et al.* (1998) involve six filter coefficients for each probe frequency for a total of 42 filter coefficients. For Glasberg and Moore (2000) the individual fits involve 30 filter coefficients. The simultaneous fits involve 12 filter coefficients, so in total, there is a reduction from 72 to 24 filter coefficients. The reduction in the number of coefficients was, of course, accompanied by a rise in rms error. For Baker *et al.* (1998), the reduction from

42 to 12 coefficients produced a rise from an average of 2.08 to 2.67 dB. For Glasberg and Moore (2000), the reduction from 30 to 12 coefficients produced a rise from an average of 1.60 to 2.62 dB. Nevertheless, the increases seem relatively small given the reduction in the number of free coefficients. The reason can be illustrated by comparing the asymmetry coefficients, c_1 and c_2, for the individual and simultaneous fits to the Glasberg and Moore (2000) data [Figs. 3(b) and 5(b), triangles]: The use of a regression line for c_1 values in the simultaneous fit means that the value of the coefficient at 0.25 kHz in Fig. 3(b) increases from what it was in the individual fit, while the value of the coefficient at 1.0 kHz is required to decrease. The asymmetry trade-off in the fitting process would lead the routine to compensate for the increase in c_1 at 0.25 kHz with a smaller value of c_2, and to compensate for the decrease in c_1 at 1.0 kHz with a larger value at c_2. But, this is exactly what the simultaneous fit is already imposing on c_2; the regression line for c_2 in Fig. 5(b) (dashed line) requires the coefficient in Fig. 3(b) at 0.25 kHz to decrease from what it was in the individual fit, while the coefficient at 1.0 kHz is required to increase from what it was.

V. GLOBAL FITTING OF THE COMBINED DATA SETS

The primary difference between the two simultaneous fits in Sec. IV is that the asymmetry coefficients, c_1 and c_2, have greater values in absolute terms for the Baker *et al.* (1998) fit. It seems unlikely that the listeners in the Baker *et al.* study really have much more asymmetric passive basilar membranes than those of the listeners in the Glasberg and Moore (2000) study, *and at the same time,* compression mechanisms with greater dynamic range. It seems much more likely that this is another example of the asymmetry trade-off in the fitting process. This suggests that the best estimate of the compressive GC filter parameters for the population of normal-hearing listeners would be provided by fitting all of the thresholds from both studies simultaneously. The fit was performed with the same procedure as for the simultaneous fits; the six filter coefficients were represented by regression lines with two coefficients each, while K and P_0 were represented by regression parabolas with three coefficients each.

The families of composite GC filters are presented in Fig. 6(a) for all seven probe frequencies in the same format as Fig. 2. The asymmetry of the compressive GC is uniform across probe frequency on this ERBrate frequency scale. The maximum gain increases with probe frequency from about 20 dB at 0.25 kHz to about 30 dB at 2.0 kHz, where it remains thereafter. The passive GC filters and HP-AFs are presented in Fig. 6(b). The passive GC filters are asymmetric and uniform across probe frequency on this scale; the compression range of the HP-AF is relatively large and grows slowly with probe frequency. The rms error increased from an average of 2.67 to 3.59 dB, but this seems reasonable given that the thresholds come from two separate studies and a very wide range of probe frequencies and levels.

The regression lines for the global fit are presented in Fig. 7. Combining the two data sets has resulted in a reduction of the slopes for b_1, c_1, and b_2; in each case, the

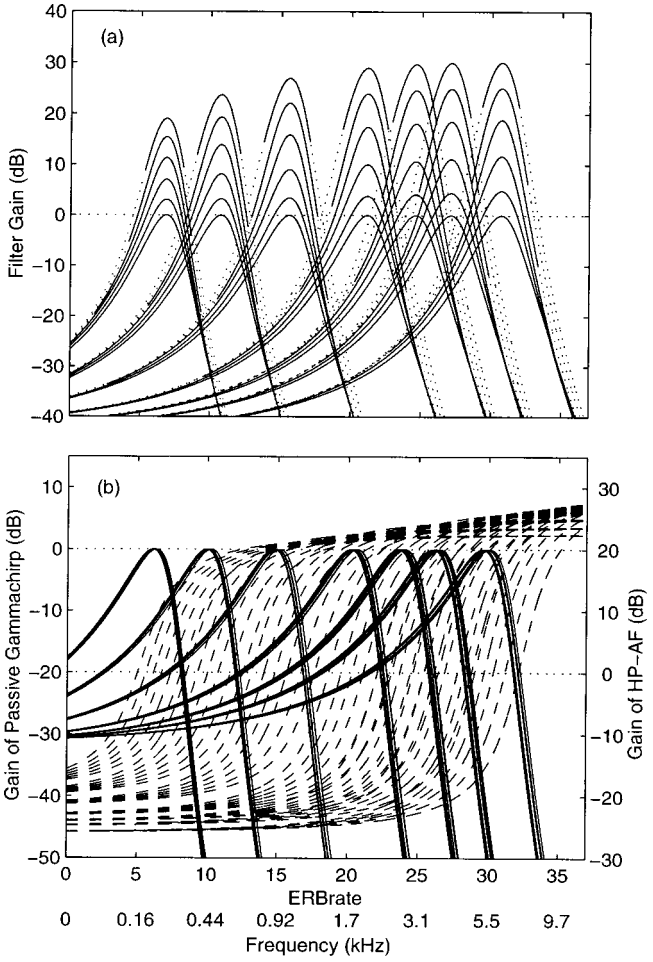


FIG. 6. Families of compressive gammachirp filters (a) and component filters (b) for the combined data of Glasberg and Moore (2000) and Baker *et al.* (1998) at all probe frequencies from 0.25 to 6.0 kHz. This global fit involved 12 regression coefficients, an intercept and slope for each of the six filter coefficients. The six curves within each set of compressive GC filters (a) and HP-AFs (b) show how the functions change as probe level increases from 30 to 80 dB in 10-dB steps.

function is now flatter than the corresponding function from either of the simultaneous fits. The slope of the regression line for c_2 has increased slightly, while the slopes for $f_{rat}^{(0)}$ and $f_{rat}^{(1)}$ are essentially the same. The slopes were so shallow as to prompt us to consider whether there was really any need for the filter coefficients to vary with probe frequency, when written as a function of ERBrate. It was clear that there was a trade-off between $f_{rat}^{(0)}$ and $f_{rat}^{(1)}$ in the fitting process, which meant that the rms error might not rise unduly if we also restricted these filter coefficients to one regression coefficient each.

These observations prompted us to perform a sequence of global fits to examine how the rms error of the filter system would increase as it was simplified. The regression coefficients for the sequence of global fits are presented in Table I along with the rms error values. The left half of the table presents values of the regression functions near their midpoint, specifically at 1.0 kHz; the right half presents the corresponding slope values in terms of the normalized ERBrate, E_f, in Eq. (9). The central column shows the rms error in dB. The first row shows the values for the global fit pre-

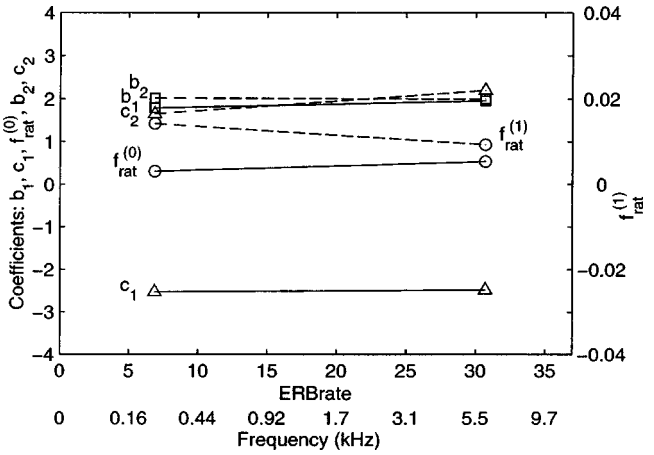


FIG. 7. Regression lines for the filter coefficients of the global fit to the combined data of Baker *et al.* (1998) and Glasberg and Moore (2000) as a function of probe frequency on an ERBrate scale. The symbols on the regression lines are not data points; they are simply there to indicate which filter coefficient is represented.

sented in Fig. 6 which has 12 filter coefficients. Since the regression functions for b_1, c_1, and b_2 were essentially flat (Fig. 7), we began by setting these three slopes to zero. The results for this nine-coefficient fit are presented in the second row, and they show that there is essentially no effect either on the rms error or the coefficient values. We then performed a pair of eight-coefficient fits, in each case setting the slope of one of the center frequency coefficients to zero ($f_{rat}^{(0)}$ or $f_{rat}^{(1)}$). The results are presented in the third and fourth rows of Table I. In both cases, setting one slope to zero resulted in a reduction of the other slope to a value near zero. The changes in the other coefficients are negligible and the increase in the rms error was very small, indicating that there was an acceptable trade-off between $f_{rat}^{(0)}$ and $f_{rat}^{(1)}$ in the region where both slopes were zero. This was confirmed by performing a seven-coefficient fit, in which the only coefficient with a nonzero slope was c_2, and showing that there was no further increase in the rms error (row 5). A plot of the families of filters for the seven-coefficient fit confirmed that they were very similar to those shown in Fig. 6.

At this point, the only remaining filter coefficient with a nonzero slope was c_2, indicating that most of the parameters of the physiological filtering system are probably proportional to frequency on an ERBrate scale. Moreover, the slope of the regression line for c_2 was also relatively shallow (0.381). Accordingly, we performed one last fit in which this slope was also set to zero. The results are presented in the bottom row of Table I. It shows that the intercept value of c_2 increases a little from 1.86 to 2.20, and there are small adjustments in the other intercept coefficients as well, but again, they are so small as to have essentially no effect on the filter shapes as presented in Fig. 6 and the increase in rms error is negligible.

In summary, this analysis indicates that it is reasonable to assume that variability in the filter parameters across frequency is effectively proportional to frequency on an ERBrate scale. There are small differences in the coefficient values as the number of filter coefficients is reduced from 12 to 6, but most of them are associated with trade-offs that have only very minor consequences for the rms error. As a result, the families of compressive gammachirp filters produced by the six-coefficient fit are very similar to those shown for the 12-coefficient fit in Fig. 6 and the corresponding plot for the six-coefficient fit is omitted for brevity.

A. Properties of the composite gammachirp filters

The *equivalent rectangular bandwidth* (ERB) of the compressive gammachirp filter is plotted as a function of probe frequency in Fig. 8 for the 12-coefficient fit; the parameter is probe level and the figure is restricted to three levels for clarity (30, 50, and 70 dB). The figure shows that the bandwidth increases slowly with level as would be expected. The dotted line with asterisks shows the ERB of the roex auditory filter at moderate stimulus levels (Glasberg and Moore, 1990); the bandwidth functions from the global fit run parallel to the ERB function, indicating that the bandwidth of the compressive GC is proportional to that of the roex; however, the compressive GC is about 1.5 times wider than the roex auditory filter. This is because, on the low-frequency side of the filter, the ratio of the slope of the passband to the slope of the tail is greater for the compressive GC. With regard to goodness of fit, this is not a crucial distinction, inasmuch as the rms errors for the best fitting roex filters obtained by Baker *et al.* and Glasberg and Moore are about the same as those obtained with the gammachirp filter. It does, however, mean that the bandwidths obtained with the

TABLE I. Filter coefficients and rms error values for six global fits of the compressive gammachirp filter to the combined data of Baker *et al.* (1998) and Glasberg and Moore (2000). Each filter coefficient was represented by a linear regression line as a function of the normalized ERB value, E_f, in Eq. (9). The first column shows the total number of filter coefficients. The left half of the table presents values of the regression functions near their midpoint, specifically the intercept at 1.0 kHz; the right half presents the corresponding slope values. The central column shows the rms error in dB. With regard to the slope coefficients, a "0" means that we set the corresponding value to zero; the "0.000" indicates that the fitting process chose a very small positive value in this case.

No.	Intercept coefficients							Slope coefficients					
coeff.	b_1	c_1	$f_{rat}^{(0)}$	$f_{rat}^{(1)}$	b_2	c_2	rms	b_1	c_1	$f_{rat}^{(0)}$	$f_{rat}^{(1)}$	b_2	c_2
12	1.85	−2.51	0.391	0.0124	2.01	1.85	3.59	0.111	0.03	0.151	−0.003	−0.012	0.359
9	1.85	−2.51	0.385	0.0125	2.02	1.85	3.62	0	0	0.149	−0.003	0	0.364
8	1.85	−2.51	0.432	0.0116	2.12	1.87	3.64	0	0	0.001	0	0	0.388
8	1.85	−2.51	0.435	0.0115	2.11	1.86	3.64	0	0	0	0.000	0	0.382
7	1.85	−2.51	0.435	0.0115	2.11	1.86	3.64	0	0	0	0	0	0.381
6	1.81	−2.96	0.466	0.0109	2.17	2.20	3.71	0	0	0	0	0	0

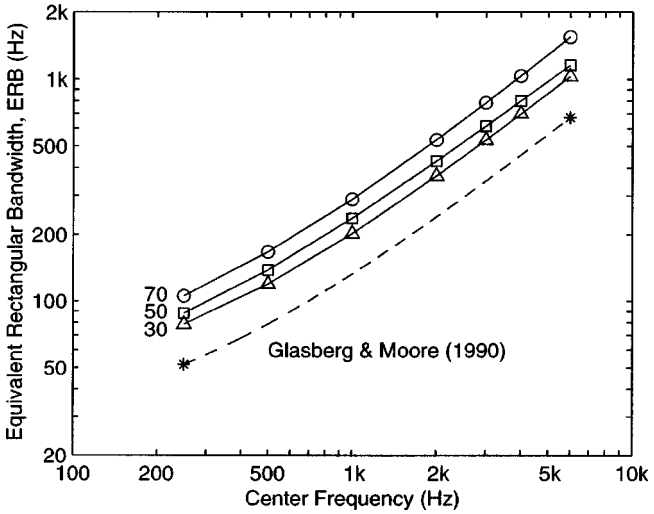


FIG. 8. Equivalent rectangular bandwidth, ERB, of the compressive gammachirp filter (solid lines) as a function of probe frequency on a log-frequency scale for the global fit with 12 regression coefficients. The parameter is probe level (30, 50, and 70 dB); the other levels are omitted for clarity. The dashed line shows the ERB of the roex filter.

gammachirp filter are more compatible with those reported by Houtgast (1977) using a simultaneous, rippled-noise masker.

The *regression parabola* obtained with the 12-coefficient, global fit for the efficiency constant, *K*, is presented in Fig. 9 (plus signs). The form agrees with that obtained by Glasberg and Moore (2000, Tables II and IV) and Baker *et al.* (1998) with the roex filter shape, in the sense that the largest values occur at high and low probe frequencies and the smallest values occur in the 1–3-kHz range. The *K* values for the compressive gammachirp are consistently smaller than those for the roex filter, which follows directly from the fact that the bandwidth of the gammachirp is greater than that of the roex. Numerically, *K* is the difference, in dB, between the average of the threshold data and the average of the thresholds predicted by the power-spectrum model, so if the integral of the filter increases, *K* must decrease for a given set of threshold data. The *K* value obtained with the compressive gammachirp is typically 2 dB more negative

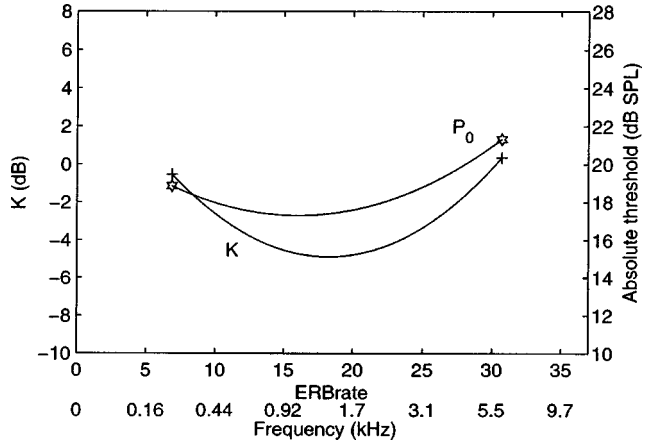


FIG. 9. The nonfilter coefficients, *K* and P_0, plotted as a function of probe frequency on the ERBrate scale for the 12-coefficient, global fit to the combined data of Baker *et al.* (1998) and Glasberg and Moore (2000).

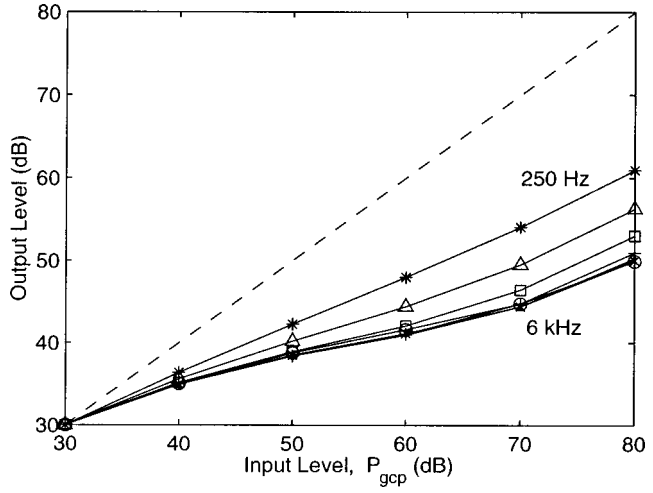


FIG. 10. Input/output functions for the global fit of the compressive gammachirp filter with 12 filter coefficients. The parameter is probe frequency from 0.25 to 6.0 kHz; the curves for 2.0–6.0 kHz are superimposed.

than that obtained with the roex, which agrees with the finding that the ERB of the gammachirp filter is about 1.5 times larger than that of the roex filter. The regression parabola for P_0 (stars in Fig. 9) is in good agreement with the P_0 results obtained by Glasberg and Moore (2000) and Baker *et al.* (1998) with the roex filter. In this case, the absolute values are also in the same range as the values reported by Glasberg and Moore (2000) for their own data and those of Baker *et al.* (1998).

In the physiological literature, *compression* is described in terms of the slope of the input–output (I/O) function of the basilar membrane. The analogous measure in psychophysical terms is the I/O function of the auditory filter, and the I/O functions obtained from the 12-coefficient, global fit are shown in Fig. 10, superimposed for all probe frequencies to facilitate comparison. The slope of the compression function gets shallower as probe frequency increases from 0.25 to 2.0 kHz, and the compression range increases from about 20 to 30 dB over the same range of probe frequencies. The slope values for the 12-coefficient, global fit are presented in Table II along with values reported by Glasberg and Moore (2000) with the roex auditory filter for their own data and those of Baker *et al.* (1998). The slope values for the roex filter were calculated from Eq. (1) of Glasberg and Moore (2000) and the $G_{\max}$ values in their Tables II and IV. The values for the compressive gammachirp agree very well with those reported for Baker *et al.* (1998); they are somewhat below the values for Glasberg and Moore (2000) for reasons that are not entirely clear.

The compression observed in these simultaneous masking studies is considerably less than that observed either in forward-masking studies (e.g., Oxenham and Plack, 1997) or in physiological studies of the basilar membrane (e.g., Ruggero *et al.*, 1992). Glasberg and Moore (2000) pointed out that the discrepancy is partially due to suppression of the signal by the masker; the masker increases the total energy of the input and thus it reduces the gain applied to the signal. In the compressive GC filter system, the filter shape is determined by the level of the signal plus masker at the output of

TABLE II. Slope values for the input/output functions of the roex filter and the compressive gammachrip filter for probe frequencies from 0.25 to 6.0 kHz. The slope of the roex filter was calculated from Eq. (1) and the G_{max} values in Tables II–IV of Glasberg and Moore (2000). The slope is calculated over the input range 30–80 dB.

Probe frequency (kHz)	0.25	0.50	1.0	2.0	3.0	4.0	6.0
roex (Table II of Glasberg and Moore, 2000)	0.73	0.70	0.39	0.56	···	0.57	···
roex (Table IV of Rosen, Baker, and Darling, 1998)	0.51	0.50	0.45	0.44	0.37	0.39	0.36
compressive GC (global fit)	0.61	0.51	0.43	0.39	0.38	0.37	0.37

the linear, passive GC filter (see Appendix B), and as a result, the compressive GC does exhibit what is usually referred to as ''two-tone suppression.'' Specifically, when a suppressor tone is introduced near a probe tone, the output of the passive GC increases and the gain of the compressive GC decreases, so the level of the probe tone at the output of the compressive GC decreases. It is also the case that the effect of suppression would be less in forward masking. In order to extend the compressive GC filter system to forward masking, however, it would be necessary to specify the temporal dynamics of the release from suppression when the masker goes off, and then estimate the relevant parameters by fitting the filter system to a large set of forward-masking data. Such an extension is beyond the scope of the current paper.

VI. SUMMARY AND CONCLUSIONS

The compressive gammachirp (GC) auditory filter is a cascade of a passive GC filter representing the action of the basilar membrane and a high-pass asymmetry function (HP-AF) representing the action of the active mechanism in the cochlea. Both the passive-GC filter and HP-AF have two parameters, one for bandwidth and one for asymmetry, and there is a fifth parameter for the frequency position of the HP-AF *relative to* the peak frequency of the passive GC. This relative frequency parameter changes with level, with the result that both the shape and the gain of the compressive GC change with level.

The filter system was fitted to two large sets of masked thresholds (Baker *et al.*, 1998; Glasberg and Moore, 2000) to estimate the parameter values of the compressive GC for the human auditory system. The threshold data spanned virtually the entire range of levels and center frequencies encountered in everyday human hearing. A sequence of fits revealed that, provided the parameters were expressed as a function of ERBrate, an excellent fit could be obtained throughout the entire domain with just one constant for each of the six filter coefficients. If we assume that the physiological parameters of the filtering system in the cochlea are uniform across the place dimension, then there may be a very simple relationship between those physiological parameters and those of the compressive GC filter system.

The gain of the filter increases as stimulus level decreases at all center frequencies. The maximum gain increases from about 20 dB at 0.25 kHz to about 30 dB at 2.0 kHz, and thereafter remains at about 30 dB. The equivalent rectangular bandwidth of the compressive GC filter is about 1.5 times that of the roex auditory filter at moderate levels, and so the bandwidth of the compressive GC filter is almost 20% of its center frequency.

ACKNOWLEDGMENTS

Brian Glasberg and Stuart Rosen kindly provided convenient files with the threshold data from Glasberg and Moore (2000) and Baker *et al.* (1998). Brian Glasberg and Brian Moore helped us understand the value of the middle-ear transforms applied to these data sets. We would also like to thank Chris Plack and Walt Jesteadt for helpful comments during the review process. The research was largely performed while the second author was a visiting scientist at the CNBH between September 2000 and August 2001 supported by a Grant-in-Aid for Science Research from the Japanese Ministry of Education (Research Fellowships of the JSPS for Young Scientists) and the British Council Grants for JSPS fellows. The research was also supported by the UK Medical Research Council (G9900369), CREST (Core Research for Evolutional Science and Technology) of JST (Japan Science and Technology Corporation), and a Special Coordination Fund for Promoting Science and Technology of young researchers with fixed-term appointments from the Japanese Ministry of Education.

APPENDIX A: BACKGROUND TO THE COMPRESSIVE GAMMACHIRP AUDITORY FILTER

The gammatone function was introduced by Johannesma (1972) and de Boer (1975) to describe cochlear impulse responses measured in cats with the reverse-correlation, or ''revcor'' technique. The envelope of the impulse response was approximated with the *gamma* distribution from statistics; the carrier, or fine structure, of the impulse response was a sinusoid, or *tone*, at the center frequency of the filter (Johannesma, 1972; de Boer, 1975). Subsequently, Schofield (1985) reported that the magnitude response of this gammatone filter could explain the masking data gathered by Patterson (1976) to derive the magnitude response, or ''shape,'' of the auditory filter psychophysically. Patterson *et al.* (1987) then showed that the magnitude response of the gammatone filter was very similar to that of the rounded exponential, or *roex*, auditory filter which had previously been used to explain human masking data by a number of modelers (for a review, see Patterson and Moore, 1986). In essence, this meant that the gammatone auditory filter could be expected to produce a reasonable time-domain simulation of cochlear

filtering in humans, and this led to the development of a succession of gammatone auditory filterbanks to simulate cochlear filtering and study the effects of phase locking on auditory perception (e.g., Meddis and Hewitt, 1991; Patterson *et al.*, 1992; Slaney, 1993; Cooke, 1993; and Patterson *et al.*, 1995). This paper describes recent attempts to extend the dynamic range and frequency domain of the gammatone filterbank using a compressive gammachirp auditory filter.

1. Level-dependent versions of the gammatone auditory filter

The original implementations of the gammatone auditory filterbank were linear and the passband of the gammatone filter was essentially symmetric on a linear frequency scale, and so they were limited to applications involving moderate stimulus levels where the auditory filter is roughly symmetric on a linear frequency scale. The masking patterns produced by narrow-band noises (e.g., Egan and Hake, 1950) showed that the auditory filter was asymmetric at high stimulus levels, with the low-frequency side shallower than the high-frequency side. The asymmetry observed in masking patterns overestimates the asymmetry of the auditory filter (Patterson, 1974); thresholds above the masker frequency are higher than those below the masker partly just because the auditory filter is broader at frequencies above the masker. Nevertheless, the auditory filter is asymmetric at high levels and so an ''asymmetric'' notched-noise method (Patterson and Nimmo-Smith, 1980) was introduced to measure the asymmetry of the roex auditory filter at relatively high stimulus levels (e.g., Lutfi and Patterson, 1984; Moore *et al.*, 1990; Rosen and Baker, 1994). The question, then, was how to introduce filter asymmetry within the gammatone framework and so produce a time-domain, level-dependent filterbank.

Physiological measurements of basilar-membrane motion confirmed that frequency selectivity was asymmetric at high stimulus levels (e.g., Pickles, 1988; Ruggero, 1992), and several level-dependent versions of the gammatone filter have recently been developed to explain physiological data. Carney (1993) paired a linear gammatone filter with a parallel nonlinear gammatone and used the system to explain revcor data. The linear gammatone had a fixed, wide bandwidth which was intended to simulate the high-level response of the passive basilar membrane; the level-dependent gammatone had a narrow bandwidth which was used to introduce a narrow passband into the composite filter at lower stimulus levels. It remained the case, however, that the magnitude response of the composite filter was largely symmetric in frequency. The latest version of this physiological gammatone filterbank is described in Zhang *et al.* (2001). Lyon (1996, 1997) developed a ''one-zero'' version of the gammatone filter to introduce asymmetry in frequency and so explain the physiological tuning curves of Ruggero (1992). Meddis *et al.* (2001) and Lopez-Poveda and Meddis (2001) developed a dual-resonance, nonlinear (DRNL) filter system that employs gammatone filters of different widths in the two routes to explain the compression and suppression observed in small mammals and some human masking data. And, Plack *et al.* (2002) have demonstrated that this DRNL system can explain the release from suppression observed in forward masking.

2. The gammachirp auditory filter

The impulse response of a filter with an asymmetric magnitude response has a frequency glide or ''chirp'' in the carrier term, and the response of the basilar membrane is known to exhibit such a chirp (Møller and Nilsson, 1979). Accordingly, Irino and Patterson (1997) introduced a chirp into the carrier term of the gammatone function to produce a ''gammachirp'' auditory filter which was then able to simulate the level-dependent asymmetry as it appears in the data of Lutfi and Patterson (1984), Moore *et al.* (1990), and Rosen and Baker (1994). However, in this ''analytic'' gammachirp, the rate of chirp at the start of the impulse response varies with level, and recent physiological data (de Boer and Nuttal, 1997, 2000; Recio *et al.*, 1998; Carney *et al.*, 1999) have shown that the rate of chirp does not vary with stimulus level. Moreover, the analytic gammachirp filter cannot account for the level-dependent gain and compression observed physiologically around the peak frequency (Pickles, 1988; Recio *et al.*, 1998). To overcome these problems, Irino and Patterson (2001) modified the architecture of the gammachirp filter to produce explicit changes in the gain and compression of the filter with stimulus level in the region of the passband of the filter, while at the same time ensuring that the form of the chirp did not change with level, as required by the physiological data. They showed that this ''compressive'' gammachirp auditory filter could explain both the chirp in the physiological revcor data of Carney *et al.* (1999) and the nonlinearities in the human masking data of Rosen and Baker (1994). It was at this point that the two new studies on human masking by Baker *et al.* (1998) and Glasberg and Moore (2000) were published.

It is currently very difficult to compare the advantages and disadvantages of the different descendants of the gammatone filterbank because of the fundamental differences in their architectures. The gammachirp is a cascade filter system; the DRNL is a parallel filter system, and Carney's model is a parallel filter system in which one channel is a cascade filter system. It is also the case that the physiological models have a large number of parameters, which makes it difficult to do global fits and assess the variation of parameter values with frequency quantitatively as in the current study. By the same token, the physiological models are more flexible and so could undoubtedly describe some data with more accuracy than the compressive gammachirp filter.

APPENDIX B: CALCULATION OF THE COMPRESSIVE GAMMACHIRP AUDITORY FILTER

The method used to calculate the compressive gammachirp auditory filter during the fitting process is illustrated by the sequence of spectra in Fig. 11. A 1000-Hz probe and an asymmetric notched noise are shown in Fig. 11(a); the beads on the probe line are intended to indicate the different probe levels in the experiment; the bars on the masker and the vertical arrow indicate that the masker level was varied to determine threshold for each probe level. This panel also

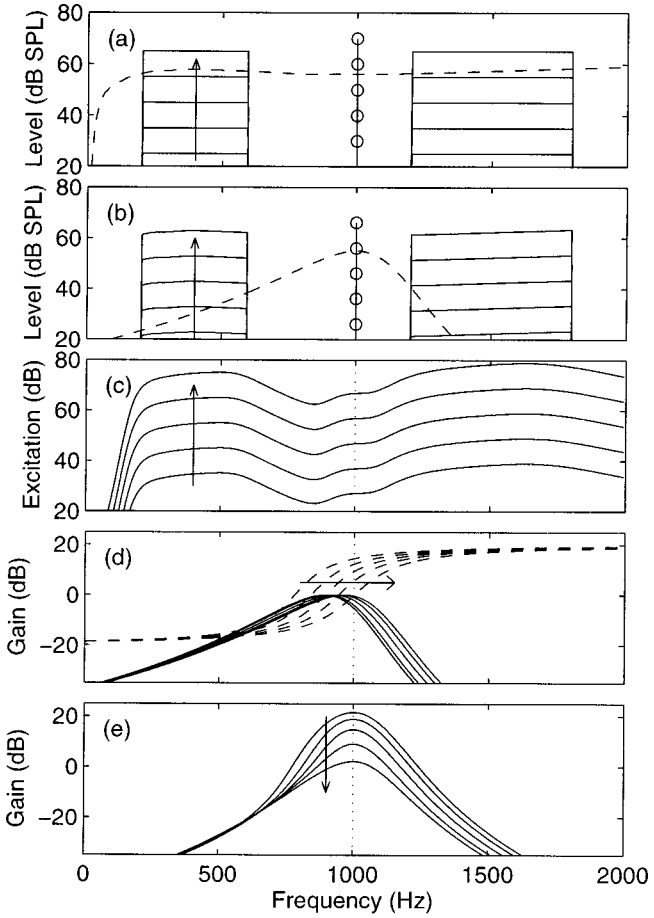


FIG. 11. A schematic of the procedure for calculating a family of level-dependent, compressive gammachirp filters with the same peak frequency as that of a given passive gammachirp.

shows the middle-ear transfer function (dashed line) used to convert the stimulus spectrum at the input to the ear canal into the spectrum at the input to the cochlea, which is shown in Fig. 11(b). The passive gammachirp (GC) filter [Fig. 11(b)] is convolved with each stimulus spectrum to simulate the excitation patterns [Fig. 11(c)] that the stimuli would produce for each probe level. These functions show the output level, P_{gcp}, of the passive GC as a function of the peak frequency of the passive GC, for these stimulus conditions.

It is assumed that the level, P_{gcp}, at the output of a particular passive GC filter determines the parameters of the compressive GC with the same peak frequency, and a typical family of compressive GC filters, for the peak frequency 1000 Hz, is presented in Fig. 11(e). These filters were produced by cascading passive GCs in the region of 1000 Hz with the HP-AFs in Fig. 11(d). The procedure is as follows: The frequency parameter, f_{rat}, is calculated from P_{gcp} as shown in Eq. (8). Then, for each compressive GC, we calculate the peak frequency of the corresponding passive GC. The remaining parameters are the same as for the passive GC filter in Fig. 11(b). Finally, we calculate the center frequency of each HP-AF [Fig. 11(d)] using Eq. (8), and cascade the HP-AFs with the passive GC to produce, the compressive GC filters in Fig. 11(e). In the fitting process, this procedure is performed in an iterative loop to find the compressive GC that produces the best probe-to-masker ratio at the output of the compressive GC.

[1]There is also a large body of data on human frequency selectivity in the form of psychophysical tuning curves; however, it has proven much more difficult to derive filter shapes from these data (e.g., see Patterson and Moore, 1986, or Bacon *et al.*, 2002), and so the current paper is limited to notched-noise data.

[2]In this eight-coefficient model, the asymptotic value of the HP-AF is fixed for individual probe frequencies. It is this which produces the uniformity of the HP-AFs outside the passband. Before making this assumption, we investigated nine- and ten-coefficient models in which the parameters that control the slope and range of the HP-AF (b_2 and c_2, respectively) were allowed to vary with level, either singly (nine-coefficient models) or jointly (a ten-coefficient model). Briefly, the results showed that allowing b_2 and c_2 to vary with level produces very little improvement in the fit beyond that produced by allowing f_{rat} to vary with level. In contrast, if f_{rat} is fixed and b_2 is allowed to vary with level, the rms error increases substantially. If f_{rat} is fixed and c_2 is allowed to vary with level, a reasonable fit is achieved; however, in this case, compression is not restricted to the passband and so this model was not pursued.

Bacon, S. P., Repovsch-Duffey, J. L., and Liu, L. (**2002**). "Effects of signal delay on auditory filter shapes derived from psychophysical tuning curves and notched-noise data obtained in simultaneous masking," J. Acoust. Soc. Am. **112**, 227–237.

Baker, R. J., Rosen, S., and Darling, A. M. (**1998**). "An efficient characterisation of human auditory filtering across level and frequency that is also physiologically reasonable," in *Psychophysical and Physiological Advances in Hearing: Proceedings of the 11th International Symposium on Hearing*, edited by A. Palmer, A. Rees, Q. Summerfield, and R. Meddis (Whurr, London), pp. 81–88.

Carney, L. H. (**1993**). "A model for the response of low-frequency auditory-nerve fibers in cat," J. Acoust. Soc. Am. **93**, 401–417.

Carney, L. H., Megean, J. M., and Shekhter, I. (**1999**). "Frequency glides in the impulse responses of auditory-nerve fibers," J. Acoust. Soc. Am. **105**, 2384–2391.

Carney, L. H., and Yin, T. C. T. (**1988**). "Temporal coding of resonances by low-frequency auditory nerve fibers: Single-fiber responses and a population model," J. Neurophysiol. **60**, 1653–1677.

Cooke, M. (**1993**). *Modelling Auditory Processing and Organisation* (Cambridge University Press, Cambridge).

de Boer, E. (**1975**). "Synthetic whole-nerve action potentials for the cat," J. Acoust. Soc. Am. **58**, 1030–1045.

de Boer, E., and de Jongh, H. R. (**1978**). "On cochlear encoding: Potentialities and limitations of the reverse-correlation technique," J. Acoust. Soc. Am. **63**, 115–135.

de Boer, E., and Nuttall, A. L. (**1997**). "The mechanical waveform of the basilar membrane. I. Frequency modulations ('glides') in impulse responses and cross-correlation functions," J. Acoust. Soc. Am. **101**, 3583–3592.

de Boer, E., and Nuttall, A. L. (**2000**). "The mechanical waveform of the basilar membrane. III. Intensity effects," J. Acoust. Soc. Am. **107**, 1497–1507.

Derleth, R. P., and Dau, T. (**2000**). "On the role of envelope fluctuation processing in spectral masking," J. Acoust. Soc. Am. **108**, 285–296.

Egan, J. P., and Hake, H. W. (**1950**). "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am. **22**, 622–630.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–61.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Glasberg, B. R., Moore, B. C. J., Patterson, R. D., and Nimmo-Smith, I. (**1984**). "Dynamic range and asymmetry of the auditory filter." J. Acoust. Soc. Am. **76**, 419–427.

Glasberg, B. R., and Moore, B. C. J. (**2000**). "Frequency selectivity as a function of level and frequency measured with uniformly exciting noise," J. Acoust. Soc. Am. **108**, 2318–2328.

Houtgast, T. (**1977**). "Auditory filter characteristics derived from direct masking data and pulsation threshold data with rippled-noise masker," J. Acoust. Soc. Am. **62**, 409–415.

Irino, T., and Patterson, R. D. (**1997**). "A time-domain, level-dependent auditory filter: The gammachirp," J. Acoust. Soc. Am. **101**, 412–419.

Irino, T., and Patterson, R. D. (**2001**). ''A compressive gammachirp auditory filter for both physiological and psychophysical data,'' J. Acoust. Soc. Am. **109**, 2008–2022.
Johannesma, P. I. M. (**1972**). ''The pre-response stimulus ensemble of neurons in the cochlear nucleus,'' in *Symposium on Hearing Theory* (IPO, Eindhoven, The Netherlands), pp. 58–69.
Lopez-Poveda, E. A., and Meddis, R. (**2001**). ''A human nonlinear cochlear filterbank,'' J. Acoust. Soc. Am. **110**, 3107–3118.
Lutfi, R. A., and Patterson, R. D. (**1984**). ''On the growth of masking asymmetry with stimulus intensity,'' J. Acoust. Soc. Am. **76**, 739–745.
Lyon, R. F. (**1996**). ''The all-pole gammatone filter and auditory models,'' Forum Acusticum '96, Antwerp, Belgium.
Lyon, R. F. (**1997**). ''All-pole models of auditory filtering,'' in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, R. Lyon, G. R. Long, and P. M. Narins (World Scientific, Singapore).
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., 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.
Møller, A. R., and Nilsson, H. G. (**1979**). ''Inner ear impulse response and basilar membrane modelling,'' Acustica **41**, 258–262.
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, London).
Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1990**). ''Auditory filter shapes at low center frequencies,'' J. Acoust. Soc. Am. **88**, 132–140.
Oxenham, A., and Plack, C. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired listening,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Patterson, R. D. (**1974**). ''Auditory filter shape,'' J. Acoust. Soc. Am. **55**, 802–809.
Patterson, R. D. (**1976**). ''Auditory filter shapes derived with noise stimuli,'' J. Acoust. Soc. Am. **59**, 640–654.
Patterson, R. D., and Henning, G. B. (**1977**). ''Stimulus variability and auditory filter shape,'' J. Acoust. Soc. Am. **62**, 649–664.
Patterson, R. D., and Moore, B. C. J. (**1986**). ''Auditory filters and excitation patterns as representations of frequency resolution,'' in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 123–177.
Patterson, R. D., and Nimmo-Smith, I. (**1980**). ''Off-frequency listening and auditory-filter asymmetry,'' J. Acoust. Soc. Am. **67**, 229–245.
Patterson, R. D., Allerhand, M., and Giguère, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.
Patterson, R. D., Holdsworth, J., Nimmo-Smith, I., and Rice, P. (**1987**). SVOS Final Report: The Auditory Filterbank. APU Report 2341.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception, Proceedings of the 9th International Symposium on Hearing*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.
Pickles, J. O. (**1988**). *An Introduction to the Physiology of Hearing* (Academic, London).
Plack, C. J., Oxenham, A. J., and Drga, V. (**2002**). ''Linear and nonlinear processes in temporal Masking,'' Acta Acust. (Beijing) **88**, 348–358.
Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (**1988**). *Numerical Recipes in C* (Cambridge University Press, Cambridge).
Recio, A. R., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). ''Basilar-membrane response to clicks at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **103**, 1972–1989.
Rosen, S., and Baker, R. J. (**1994**). ''Characterizing auditory filter nonlinearity,'' Hear. Res. **73**, 231–243.
Rosen, S., Baker, R. J., and Darling, A. (**1998**). ''Auditory filter nonlinearity at 2 kHz in normal-hearing listeners,'' J. Acoust. Soc. Am. **103**, 2539–2550.
Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Curr. Opin. Neurobiol. **2**, 449–456.
Ruggero, M. A., Rich, N. C., Recio, A. R., Narayan, S. S., and Robles, L. (**1992**). ''Basilar-membrane response to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Schofield, D. (**1985**). ''Visualizations of speech based on a model of the peripheral auditory system,'' NPL Report DITC 62/85.
Slaney, M. (**1993**). ''An efficient implementation of the Patterson–Holdsworth auditory filterbank,'' Apple Computer Technical Report #35.
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.

Informational masking and musical training

Andrew J. Oxenham[a)]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and Harvard-MIT Division of Health Sciences and Technology, Speech and Hearing Bioscience and Technology Program, Cambridge, Massachusetts 02139

Brian J. Fligor, Christine R. Mason, and Gerald Kidd, Jr.
Hearing Research Center and Programs in Communication Disorders, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 27 November 2002; revised 23 May 2003; accepted 16 June 2003)

The relationship between musical training and informational masking was studied for 24 young adult listeners with normal hearing. The listeners were divided into two groups based on musical training. In one group, the listeners had little or no musical training; the other group was comprised of highly trained, currently active musicians. The hypothesis was that musicians may be less susceptible to informational masking, which is thought to reflect central, rather than peripheral, limitations on the processing of sound. Masked thresholds were measured in two conditions, similar to those used by Kidd *et al.* [J. Acoust. Soc. Am. **95**, 3475–3480 (1994)]. In both conditions the signal was comprised of a series of repeated tone bursts at 1 kHz. The masker was comprised of a series of multitone bursts, gated with the signal. In one condition the frequencies of the masker were selected randomly for each burst; in the other condition the masker frequencies were selected randomly for the first burst of each interval and then remained constant throughout the interval. The difference in thresholds between the two conditions was taken as a measure of informational masking. Frequency selectivity, using the notched-noise method, was also estimated in the two groups. The results showed no difference in frequency selectivity between the two groups, but showed a large and significant difference in the amount of informational masking between musically trained and untrained listeners. This informational masking task, which requires no knowledge specific to musical training (such as note or interval names) and is generally not susceptible to systematic short- or medium-term training effects, may provide a basis for further studies of analytic listening abilities in different populations.
[DOI: 10.1121/1.1598197]

PACS numbers: 43.66.Dc, 43.75.Cd [MRL]

I. INTRODUCTION

The ability to distinguish, or ''hear out,'' a sequence of tones, such as a melodic line, in the presence of other tones is fundamental to the appreciation of most forms of music. When individual tones are not perceived due to the presence of other sounds, they are said to be masked. The most commonly studied form of masking, sometimes referred to as ''energetic masking'' (Pollack, 1975; Leek *et al.*, 1991; Kidd *et al.*, 1994), is thought to be determined primarily by the frequency selectivity of the peripheral auditory system: when the masker and signal are sufficiently close in frequency, and when the masker is sufficiently intense, the peripheral neural representation of the masker dominates that of the signal to such an extent that no subsequent cognitive strategies are sufficient to extract the signal.

Masked thresholds can also be influenced by attention, listener expectations, and uncertainty about the signal's characteristics. For instance, if the listener is presented with a background of white noise and, through experimental manipulations or instructions, is expecting a tone of a particular frequency or duration to be presented, tones differing in some way from expectations are often more poorly detected (Greenberg and Larkin, 1968; Scharf *et al.*, 1987; Hafter *et al.*, 1993; Wright and Dai, 1994).

Finally, masked thresholds in situations involving *masker* uncertainty are often much higher than predicted by energetic masking alone. For instance, if the masker consists of a number of tones having frequencies that are changed randomly from presentation to presentation, listeners often have great difficulty in detecting a signal of a fixed and known frequency, even if the masker frequencies are always far removed from that of the signal. Masking that cannot be explained in terms of peripheral frequency selectivity is often referred to as ''informational masking'' (Pollack, 1975; Watson, 1987; Neff *et al.*, 1993), and is almost certainly mediated at higher stages of perceptual processing.

Informational masking is often accompanied by large differences in performance between listeners (e.g., Lutfi *et al.*, 2003). In a comprehensive study using up to 49 listeners, Neff and Dethlefs (1995) found individual differences in masked thresholds as large as 50 dB. For detecting a 1-kHz signal in a masker comprised of ten random-frequency tones, the standard deviation around the mean was 11 dB. This is considerably larger than the standard deviations of 2 dB or less often found for the detection of the same signal in broad-

a)Electronic mail: oxenham@mit.edu

band noise. The large effects of, and differences in, informational masking persist even after extended practice (Neff and Callaghan, 1988; Neff and Dethlefs, 1995).

One obvious question is whether listeners at the extreme ends of performance in informational masking employ different listening strategies. A cursory analysis performed by Neff and Dethlefs (1995) showed that this might indeed be the case. They found that their ''best'' listeners had thresholds that seemed to be determined by energetic masking. That is, their thresholds were similar to what would have been expected from an ''optimal'' single-channel model, where only the output of the auditory filter centered on the signal frequency was analyzed. At the other extreme, their ''worst'' listeners had thresholds that could be predicted based on the level of the overall stimulus: these listeners seemed unable to ''hear out'' the signal and instead may have based their judgments on an impression of the overall loudness of the total tone complex. In summary, a distinction could be made between ''analytic'' and ''holistic'' listeners, although the population formed a continuum rather than two distinct groups.

So far, it has proved challenging to identify predictors of how listeners will perform in an informational masking task. Neff *et al.* (1996) found a significant effect of sex in their group of 49 listeners, with females tending to have somewhat higher thresholds than males. Other measures of auditory performance, such as absolute thresholds or thresholds for a tone in broadband noise, have yielded no significant correlations with thresholds in informational masking situations (Neff *et al.*, 1993; Neff and Dethlefs, 1995).

Musicianship, both active and passive, relies in part on an ability to listen analytically. Thus, a reasonable hypothesis is that listeners with a high level of musical training and ability (musicians) should be able to listen more analytically, and hence perform better in informational masking tasks, than listeners with no formal musical training (nonmusicians). No study has yet directly tested this hypothesis. However, other studies have indicated that nonmusicians may have more difficulty in tasks requiring analytic listening. Soderquist (1970) found that nonmusicians were less able to identify individual partials within a harmonic tone complex than were musicians. He described this difference in terms of poorer frequency selectivity on the part of nonmusicians. Fine and Moore (1993) confirmed Soderquist's finding that nonmusicians had more difficulty than musicians in identifying individual tones within a complex. However, they also measured auditory filter shapes, using the notched-noise method, whereby masked thresholds for a tone are measured in the presence of noise with a spectral notch (Patterson, 1976; Glasberg and Moore, 2000). They found no relationship between the bandwidth of the auditory filters and level of musical training. In short, while the ability to identify tones in a background of other tones might be poorer in nonmusicians, there is no evidence for poorer peripheral frequency selectivity, as originally implied by Soderquist (1970).

Another study comparing the performance of musicians and nonmusicians (Spiegel and Watson, 1984) investigated frequency discrimination. The tasks included a simple comparison of two successive tones and a more complex task, where the target tones formed part of a longer sequence of tones. The general conclusion was that the musicians' performance was initially better but that, with extended training, the performance of nonmusicians could be brought up to that of musicians. The effects of training are particularly striking in the frequency discrimination and pattern recognition tasks employed by Spiegel and Watson (1984). In contrast, practice effects in informational masking tasks appear to be much less robust. The two studies that have examined learning effects in informational masking (Neff and Callaghan, 1988; Neff and Dethlefs, 1995) both found little evidence for systematic learning effects: although certain individuals did show improvements with time, there was generally no systematic trend when examining group data.

The assumption that good performance requires analytic listening, together with the finding that practice effects tend not to be particularly large or systematic, makes informational masking an attractive task for testing analytic listening abilities. The aim of this study was to establish whether musicians do in fact exhibit less informational masking than do nonmusicians, as might be expected if musicians are able to listen more analytically. Detection thresholds were measured in an informational masking task, similar to that used by Kidd *et al.* (1994), in two groups comprising musicians and nonmusicians, with equal numbers of males and females in each group. In the same group of listeners, frequency selectivity was measured using a version of the notched-noise method (Stone *et al.*, 1992; Glasberg and Moore, 2000). This enabled us to test for any differences in peripheral frequency selectivity in our two groups. Based on the data of Fine and Moore (1993), none was expected.

II. METHODS

A. Stimuli

All sounds were computer generated at a rate of 20 kHz, were played through 16-bit digital-to-analog converters (Tucker-Davis Technology), and were then low-pass filtered at 7500 Hz. The signal was a single 200-ms (total duration) burst of a 1000-Hz tone for the noise-masker conditions, or a sequence of 60-ms (total duration) 1000-Hz tone bursts for the multitone masking conditions. In both cases, the signals were gated on and off with 10-ms raised-cosine ramps. In all masking conditions, the signal was presented at a fixed level of 20 dB above absolute threshold, adjusted for each subject individually.

The tonal maskers were comprised of a series of random-frequency multitone complexes. The multitone complexes were played in a sequence of eight contiguous bursts having rise/steady-state/decay times of 10/40/10 ms, for a total duration of 480 ms. The signal, when present, was gated on and off synchronously with the masker bursts. There were two versions of the multitone masker that differed in the way the frequencies of the tones were randomized. For one masker, referred to as ''multiple-bursts same'' (MBS), the frequencies of the eight masker tones in the first burst were randomly drawn from a uniform distribution of frequencies, on a logarithmic scale, ranging from 200–5000 Hz excluding

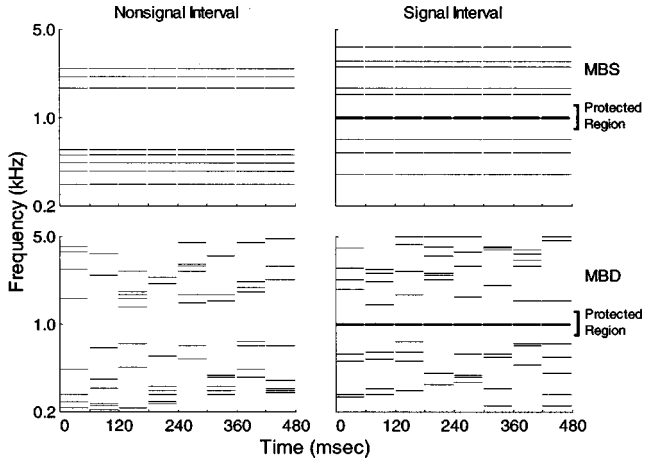


FIG. 1. Schematic diagram of the stimuli used in the random-frequency multitone masker conditions. The left panels show typical masker-alone samples that might occur in the nonsignal interval of a trial, while the right panels show typical signal-plus-masker samples that might occur in the signal interval of a trial. The signal is shown with heavy lines. The upper row illustrates the ''multiple-bursts same'' (MBS) masker while the lower row illustrates the ''multiple-bursts different'' (MBD) masker (see the text for details).

a ''protected region'' with a bandwidth of 400 Hz centered geometrically around the 1000-Hz signal frequency. This protected region is designed to reduce the amount of energetic masking produced when the masker and signal frequencies are close together. For the MBS masker, the frequencies chosen in the first burst in a sequence were repeated in the seven subsequent masker bursts for that sequence. A different set of random frequencies was chosen for each sequence. The other multitone masker is referred to as ''multiple-bursts different'' (MBD). For the MBD masker, the frequencies for every burst in a sequence were drawn randomly in exactly the same way as for the first burst of MBS. Thus, each sequence of MBD contained eight bursts that were different random frequency draws. The stimulus configuration was similar to that used in earlier studies (Kidd *et al.*, 1994; 2002b) and has been shown to provide a robust informational masking effect in the MBS condition, as well as a large masking difference between the MBS and MBD conditions. Figure 1 illustrates these two multitone maskers schematically in sound spectrogram form.

The figure shows typical draws of both MBS and MBD maskers with the right column indicating a masker plus signal and the left column illustrating the masker alone. Also shown is the protected region around the signal frequency where masker tones were not permitted to fall. The levels of the tones within a masker draw were equal.

A Gaussian noise was used to estimate peripheral frequency selectivity. The noise had a flat spectrum within the bandpass region and was filtered in many conditions to introduce a spectral notch. These conditions were used to estimate the characteristics of the auditory filter containing the signal and the processing efficiency of the listener for the tone-in-noise detection task (Patterson, 1976; Glasberg and Moore, 2000). The noise was presented as a single continuous burst having a duration of 300 ms. The 200-ms, 1000-Hz signal was temporally centered in the noise. The five notch widths were given by: 0.0 and 0.0 (no notch); 0.2 and 0.2; 0.4 and 0.4; 0.2 and 0.4; and 0.4 and 0.2, expressed as the difference between signal frequency and notch edge frequency, divided by signal frequency. This abbreviated version of the notched-noise test has been found in the past to provide reliable results when deriving auditory filter shapes (Stone *et al.*, 1992). The noise had a bandwidth of 400 Hz on either side of the notch.

B. Subjects

A total of 24 adult subjects with normal hearing (thresholds of 15 dB HL or less at octave frequencies between 250 and 8000 Hz) served as listeners in these experiments. Their ages ranged from 19 to 47 years (mean age 24.7; median 22.5). Twelve subjects were trained musicians and 12 subjects identified themselves as nonmusicians. In both groups males and females were equally represented. The mean age of the musicians was 28.3 years (s.d. 8.3) and mean age of the nonmusicians was 21.1 years (s.d. 2.2). The mean absolute threshold for the multiple-burst signal was 3.2 dB SPL (s.d. 5.2 dB) and 3.3 dB SPL (s.d. 5.7 dB) for the 200-ms signal used in the noise-masking conditions. There were no significant differences in mean absolute threshold between the musician and nonmusician groups for either signal $[|t(22)|<1.6;\ p>0.1]$.

The selection criteria for inclusion in the musician group were as follows: first, all subjects had musical training at the college level in addition to 2 or more years of private lessons (virtually all reported beginning their musical training formally or informally as children). Nine were currently students in college-level music programs and two were graduates of music programs. Two reported holding graduate degrees in music. One other subject studied at the college level for 3 years and then became (and remains) a professional musician engaged in recording and performing. All currently play musical instruments regularly and had at least 2 semesters of formal ear training or, in one case, the subject tested out of the college ear-training requirement. In addition, all subjects were able to achieve 90% or higher accuracy on a relative pitch test. In the relative pitch test, the listener was given a pure tone of a frequency corresponding to the musical note ''A'' (440 Hz). Following that tone, another tone was presented having a frequency equaling that of one of the 12 notes on the semitone scale beginning with and extending above middle ''C'' (261.63 to 493.88 Hz) and the listener was asked to name the musical interval. The listener made 20 such judgments. For the nonmusicians, nine of the subjects reported that they had never had any musical training whatsoever. The other three subjects reported minimal experience attempting to learn to play musical instruments as children—including some private lessons or lessons in elementary school—but did not continue to play the instrument or take further lessons past the age of 10 years. None of the nonmusicians currently played any musical instrument and only one (not one who had taken lessons as a child) indicated some knowledge of how to read music (the subject gained some experience following written music as a dancer).

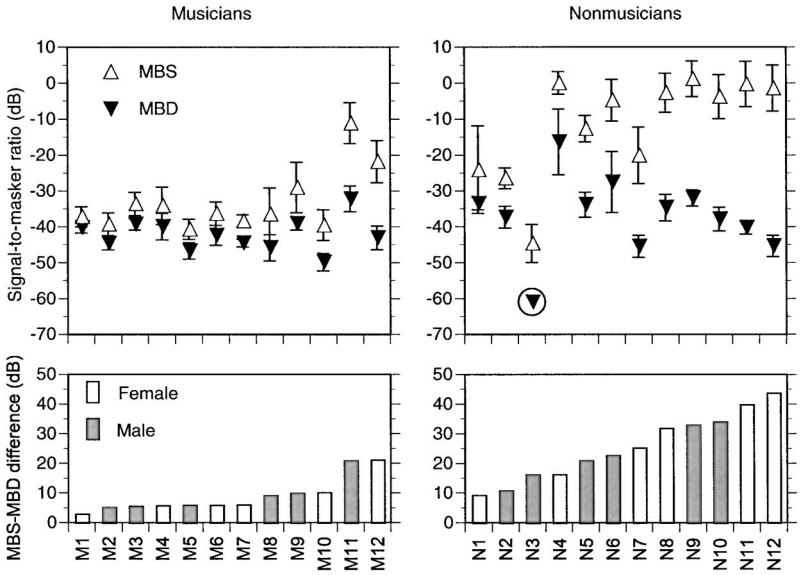


FIG. 2. Individual results from the random-frequency multitone masker conditions. The left panels show data from the musician group and the right panels show data from the nonmusician group. The upper panels show the signal-to-masker ratios at threshold in the multiple-burst same (MBS; upward-pointing open triangles) and the multiple-burst different (MBD; downward-pointing filled triangles) conditions. The symbol in a circle represents a point that could not be measured. Error bars denote ± 1 s.d. of the mean. The lower panels show the difference in masking between the two conditions. The subjects are arbitrarily ordered according to the magnitude of the difference in masked threshold between MBS and MBD masking conditions. Shaded and unshaded bars in the lower panels denote male and female listeners, respectively.

C. Procedures

Thresholds were measured using an adaptive 2-interval, 2-alternative forced-choice procedure that estimates the 70.7%-correct point on the psychometric function (Levitt, 1971). The level of the adaptively changing stimulus was varied initially in 6-dB steps, which was reduced to 3-dB steps after the first four reversals. There were 50 trials in each block of trials, and a threshold estimate was counted as valid only if at least nine reversals were obtained. The average of the reversals after discarding the first three or four reversals (whichever resulted in an even number) was then computed. Response feedback was provided after every trial. Initially, in each session, threshold was measured twice for the signal presented in quiet (no masker). If the two estimates differed by more than 3 dB a third estimate was obtained. After that the signal level was fixed at a level 20 dB above the average of the quiet threshold estimates (20 dB SL) for subsequent masking conditions.

In the masked conditions, the masker was varied using the same procedures as for the signal in quiet, except that the signal level was fixed and masker level was varied adaptively. The listeners were tested first on the Gaussian notched-noise masking task (except for one listener who was tested on the informational masking task first due to experimenter error) with a minimum of four estimates obtained for each notch width. After collection of those data was complete, the listeners were tested on the multitone masking conditions with the two types of maskers alternated in sets of two with at least eight estimates obtained for each type of multitone masker.

The stimuli were presented to one ear using a calibrated TDH-50 earphone in a double-walled IAC booth. The listeners were tested individually for two sessions lasting approximately 2 h each, except for one listener who required a brief third session to finish data collection. During the first session, the listeners' hearing was tested using standard pure-tone air-conduction audiometry and a questionnaire was administered on musical training and background. For the listeners in the musician group, the relative pitch test was also conducted at some point during the testing sessions.

III. RESULTS

A. Random-frequency multitone masking

For each listener in each condition, the signal-to-masker ratio at threshold was calculated by subtracting the overall masker level at threshold from the signal level (both in dB SPL), which was fixed at 20 dB SL individually for each listener, with a mean presentation level of about 23 dB SPL.[1] Because signal thresholds in quiet varied somewhat across listeners, expressing thresholds in terms of signal-to-masker ratio provided a way of comparing performance across listeners more directly. The data were first analyzed to test for learning effects across the eight runs completed by each listener in every condition. This was done by performing a within-subjects linear regression of threshold as a function of repetition number separately for the musicians and nonmusicians in the MBD and MBS conditions. For the musicians in the MBD condition, there was a significant trend for thresholds to improve with repetition number [$F(1,11)=63.6$, $p<0.001$]. However, the overall improvement was rather small, with a mean improvement of 0.5 dB per repetition. None of the other conditions showed a significant trend for threshold changes across the eight repetitions ($p>0.05$). Given that only eight runs were made, our data do not address the question of longer-term learning. However, they are generally consistent with earlier findings of rather weak learning effects in random-frequency multitone masking experiments (Neff and Callaghan, 1988; Neff and Dethlefs, 1995).

The individual mean signal-to-masker ratios at threshold are plotted in the top two panels of Fig. 2. The left panel shows data from the musicians, while the right panel shows data from the nonmusicians. The sex of individual subjects is indicated by the shading of the bars in the lower panels. Open upward-pointing triangles denote thresholds from the MBS condition; filled downward-pointing triangles denote thresholds from the MBD condition. The error bars represent

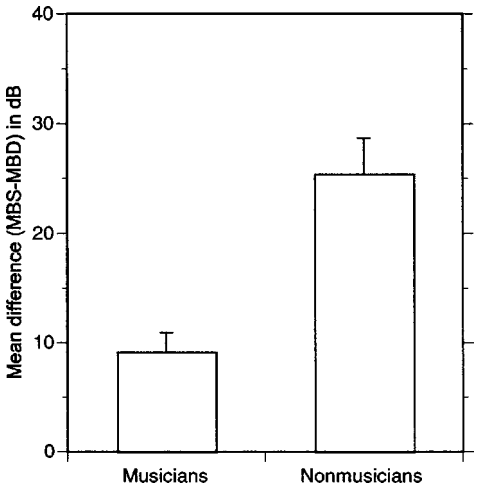


FIG. 3. Mean masking differences between the two types of random-frequency multitone maskers, MBS and MBD, for the two groups of listeners. Error bars denote +1 standard error of the mean.

±1 s.d. of the mean. Consider first the MBD condition (filled symbols). The results are reasonably similar both within and across the two groups. This impression was confirmed by a *t*-test, showing no significant difference in the mean signal-to-masker ratio for the MBD condition between the two groups [$t(22)=1.54$; $p>0.1$]. The generally good and relatively uniform performance in the MBD condition is consistent with the view that thresholds in this condition depend largely on energetic masking (Kidd *et al.*, 1994): the constant frequency of the signal repetitions assists the listener in perceptually segregating the signal from the maskers, thereby avoiding (or at least reducing) informational masking.

Consider next the MBS condition (open symbols). Thresholds are considerably higher in this condition (indicating poorer performance), as expected based on previous results (e.g., Kidd *et al.*, 1994). While there are some subjects in the nonmusician group whose performance is comparable to those in the musician group, the overall performance of the nonmusician group is considerably poorer. This impression was confirmed by a *t*-test showing a highly significant difference between the two group means for the MBS signal-to-masker ratios [$t(22)=4.44$; $p<0.001$].

If the interpretation that the MBD condition reflects primarily energetic masking is correct, then the *difference* between thresholds in the two conditions should provide a good measure of the amount of additional informational masking produced by the MBS condition. These results are shown in the lower panels of Fig. 2. Overall, the smaller differences observed in the musician group are consistent with the hypothesis, outlined in the Introduction, that musicians should exhibit less informational masking than nonmusicians. The mean differences for the two groups are shown in Fig. 3.

As mentioned in the Introduction, Neff *et al.* (1996) reported a sex effect in informational masking, with female listeners as a group exhibiting significantly higher thresholds (poorer performance) than male listeners. However, Neff *et al.* (1996) did not report on the musical abilities of their listeners. Our two groups (musicians and nonmusicians) were selected so that both sexes were equally and evenly represented. This enabled us to test both for effects of musical training and for sex. An examination of the distribution of males and females in the lower half of Fig. 2 suggests no clear effect of sex for either musicians or nonmusicians. A two-factors (musical training and sex) between-subject analysis of variance (ANOVA) was performed, with the MBS–MBD difference as the dependent variable. The results were clear: there was a significant effect of musical training [$F(1,20)=18.56$; $p<0.001$], confirming our earlier analysis, while neither the effect of sex [$F(1,20)<1$] nor the interaction between musical training and sex [$F(1,20)<1$] was significant. Sex remained a nonsignificant effect when the analysis was carried out using only the MBS or MBD signal-to-masker ratios.

Our results are not consistent with the findings of Neff *et al.* (1996), in that we find no effect of sex on performance. It is not clear what accounts for this difference. While we used a smaller number of subjects than did Neff *et al.* (24 vs 49), it is unlikely that simply adding more subjects would produce a significant result in our case, given that the effect of sex did not even approach significance. In light of the present study, one might speculate that there were more musicians among Neff *et al.*'s male listeners than among their female listeners. In any case, no hint of an effect of sex was found here, when listeners were balanced for musical training, suggesting that men are not necessarily better listeners than women.

As mentioned in the Methods section, the mean age of the musician group was somewhat higher than that of the nonmusician group. To test for any effect of age, a between-subjects analysis of covariance (ANCOVA) was performed with the MBS–MBD difference as the dependent variable, musicianship as a factor, and age as a covariate. As expected from the previous analyses, the effect of musicianship was highly significant [$F(1,21)=12.99$, $p<0.002$], but the effect of age was not ($F<1$). Thus, the difference in mean age between the two groups cannot account for the differences in informational masking.

Finally, some recent work has examined the effects of hearing impairment on informational masking (Micheyl *et al.*, 2000; Kidd *et al.*, 2002a). While all our listeners qualified as normal-hearing according to ANSI (1969) standards, there were substantial variations in absolute threshold for the signal, ranging from −9 dB to 15 dB SPL. However, absolute thresholds seemed to play no role in the amount of informational masking observed; another ANCOVA with absolute threshold as the covariate found no significant effect of absolute threshold ($F<1$). Separate analyses of musician and nonmusician data confirmed the lack of correlation between the MBS–MBD difference and absolute threshold (Pearson product-moment correlations of 0.03 and −0.23, respectively; $p>0.2$ for both).

B. Notched-noise masking: Auditory filter shapes

The individual masker thresholds from the notched-noise conditions, using the fixed 1000-Hz 200-ms signal with the level set 20 dB above individual thresholds in quiet (mean presentation level of about 23 dB SPL), were ana-

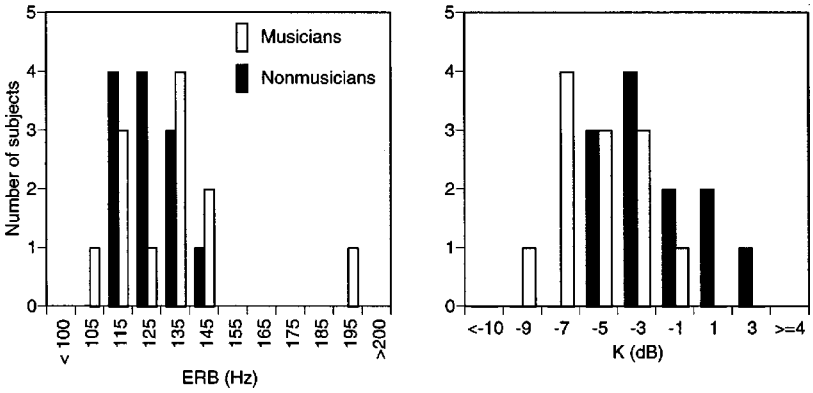


FIG. 4. Results from fitting auditory filters to the individual data from the notched-noise conditions. The data are plotted in the form of histograms, with the values on the abscissa representing the midpoint of each bin. The left panel shows estimates of the equivalent rectangular bandwidth (ERB) of the individual auditory filters, and the right panel shows estimates of detection efficiency, K.

lyzed and used to derive auditory filter shapes and detection efficiency values using the roex(pr) model of auditory filter shape (Patterson and Moore, 1986). The filter weighting function (W) is assumed to be of the form

$$W(g)=(1+p|g|)e^{-p|g|}+r, \tag{1}$$

where g is the normalized frequency distance from the filter's center frequency (distance divided by center frequency), p is a parameter defining the sharpness of filter tuning, and r is a constant, designed to limit the dynamic range of the filter. The value of p is allowed to be different below (p_l) and above (p_u) the filter's center frequency, thereby allowing asymmetry in the filter shape. The best-fitting values of p were determined using a least-squares minimization routine. The detection efficiency, K, in dB is the mean signal-to-noise ratio (in dB) across all conditions for the best-fitting filter shape. For further details see, e.g., Glasberg and Moore (1990).

The resulting filter shapes were reasonably uniform across listeners, both in terms of their asymmetry and their bandwidths. Because of this, only the equivalent rectangular bandwidths (ERBs) of the filters are discussed. The values of the ERBs and Ks derived from the individual data from notched-noise experiment are shown in Fig. 4 in the form of histograms. Bin widths of 10 Hz and 2 dB were used for the ERB and K values, respectively. Higher ERB values imply poorer frequency selectivity; higher K values imply poorer detection efficiency. Data from musicians are shown as open bars; data from the nonmusicians are shown as filled bars. Consider first the ERB values. The distributions of the two groups are rather similar, as are the mean ERB values of 134 Hz for the musicians and 126 Hz for the nonmusicians. These values are also in very good agreement with the value of 132 Hz at a center frequency of 1 kHz, given by the formula of Glasberg and Moore (1990), which was based on data from a number of earlier studies. A two-way ANOVA showed that neither the effect of sex nor musical training (or their interaction) was significant ($p>0.1$ in all cases). This confirms Fine and Moore's (1993) finding that musical training has no effect on peripheral frequency selectivity.

Consider next the values of detection efficiency, K (Fig. 4, right panel). There seems to be a slight trend for K values to be lower in musicians, indicating somewhat greater detection efficiency. This is reflected in the mean K values of -3.43 and -1.97 dB for the musicians and nonmusicians, respectively, and is also consistent with the findings of Fine and Moore (1993). However, in contrast to that earlier study, the trend observed in our data was not statistically significant. In a two-way between-subjects ANOVA, neither the effect of sex nor musical training (or their interaction) reached significance ($p>0.1$ in all cases). When sex was ignored (as in the Fine and Moore study), a simple pooled t-test comparing the two groups of musicians and nonmusicians still revealed no significant difference [$t(22)=1.49$; $p>0.1$].

The distribution of K values appears somewhat skewed towards positive values. For this reason, the analyses were repeated using the nonparametric Mann–Whitney U-test and the Kolmogorov–Smirnov Z-test (which do not assume normal distribution). Both tests also failed to show significant differences in either ERB or K between musicians and nonmusicians. A comparison of our listener groups with those of Fine and Moore (1993) shows that the two musician groups performed similarly, whereas our nonmusician group showed somewhat lower (more efficient) K values than did those of Fine and Moore. In summary, filter bandwidth and detection efficiency were not significantly related to musical training, at least in our group of 24 listeners.

IV. DISCUSSION

This study investigated the effects of musical training on informational and energetic masking using random-frequency multitone and noise maskers. In the random-frequency multitone masking condition thought to involve informational masking (MBS condition), a large and statistically significant difference in performance was found between a group of musically trained listeners (musicians) and a group of listeners with no musical training (nonmusicians). In the random-frequency multitone condition, thought to rely more on energetic masking (MBD condition), no significant difference between musicians and nonmusicians was found. The reduced susceptibility to informational masking found in musicians may reflect superior analytic listening abilities in trained musicians. Here, ''analytic listening'' refers specifically to the ability to discern or ''hear out'' a predefined partial of a complex sound. In some ways this can be considered analogous to situations in which musicians must follow individual musical ''voices'' in unfamiliar pieces, although of course the stimuli used here do not resemble any traditional musical forms, nor do the tones (pure sinusoids

with no amplitude fluctuations) resemble any physical instruments. In contrast to a previous study (Neff *et al.*, 1996), no effect of sex was found in any condition.

Peripheral frequency selectivity and detection efficiency were estimated using a version of the notched-noise technique. The data, as summarized using the ERB and detection ratio *K*, revealed no significant differences in performance between musicians and nonmusicians. This finding is consistent in part with the findings of an earlier study (Fine and Moore, 1993), although in that study detection efficiency was found to be significantly better in musicians than in nonmusicians.

The informational masking paradigm used in this study seems to tap into a genuine difference in listening abilities between people with and without a high degree of musical training. Considering the difference in thresholds between the MBS and MBD conditions (lower panels of Fig. 2 and Fig. 3), it can be seen that there is surprisingly little overlap in performance between the two groups. For instance, an attempt to discriminate musicians from nonmusicians solely by their MBS–MBD threshold difference by setting an arbitrary criterion of 10.5 dB leads to only two misidentified musicians and one misidentified nonmusician out of a total of 24. The finding of a robust difference between the groups is interesting because the task is not one in which musicians are specifically trained: the stimuli, comprising repeated bursts of random-frequency sinusoids gated in precise synchrony, are extremely unnatural. Furthermore, the task is not one that requires any explicit musical training, such as the ability to name notes or musical intervals.

In summary, we have shown a link between musical training and performance in an informational-masking task. Our study is by its nature correlational. Because of this, there is no way of telling whether it is musical training, a natural inclination towards pursuing music, or some other common factor that results in better performance in the musician group. However, the informational-masking task seems to provide a quantitative measure of analytic listening ability, without requiring knowledge specific to Western music, such as note names or music notation, or any high cognitive load, such as remembering a melodic line. The task may therefore prove to be a useful tool in future studies of analytic listening and musical ability in general.

ACKNOWLEDGMENTS

This work was supported by NIH Grants R01 DC 04545 (G.K.) and R01 DC 03909 (A.J.O.). The first version of this paper was written while the first author was a fellow at the Hanse Institute for Advanced Study in Delmenhorst, Germany. We thank Tanya Arbogast, Christine Carter, and Sally Tressler for assistance. Christophe Micheyl, Bertram Scharf, Charles Watson, and Marjorie Leek provided helpful comments on earlier versions of the paper.

[1]For one nonmusician it was not possible to measure a threshold in the MBD condition, as he could still detect the signal when the masker was at the highest permitted level of 90 dB SPL. This is shown in the upper-right panel of Fig. 2 by the symbol in a circle. For the purposes of analysis, the listener's threshold was set to the maximum allowable level, as shown by the position of the symbol.

ANSI (**1969**). ''Specifications for audiometers,'' American National Standards Institute, ANSI S3.6-1969.

Fine, P. A., and Moore, B. C. J. (**1993**). ''Frequency analysis and musical ability,'' Music Percept. **11**, 39–53.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Glasberg, B. R., and Moore, B. C. J. (**2000**). ''Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise,'' J. Acoust. Soc. Am. **108**, 2318–2328.

Greenberg, G. S., and Larkin, W. D. (**1968**). ''Frequency-response characteristics of auditory observers detecting signals of a single frequency in noise: The probe-signal method,'' J. Acoust. Soc. Am. **44**, 1513–1523.

Hafter, E. R., Schlauch, R. S., and Tang, J. (**1993**). ''Attending to auditory filters that were not stimulated directly,'' J. Acoust. Soc. Am. **94**, 743–747.

Kidd, Jr., G., Arbogast, T. L., Mason, C. R., and Walsh, M. (**2002a**). ''Informational masking in listeners with sensorineural hearing loss,'' J. Assoc. Res. Otolaryngol. **3**, 107–119.

Kidd, Jr., G., Mason, C. R., and Arbogast, T. L. (**2002b**). ''Similarity, uncertainty, and masking in the identification of nonspeech auditory pattern,'' J. Acoust. Soc. Am. **111**, 1367–1376.

Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). ''Reducing informational masking by sound segregation,'' J. Acoust. Soc. Am. **95**, 3475–3480.

Leek, M. R., Brown, M. E., and Dorman, M. F. (**1991**). ''Informational masking and auditory attention,'' Percept. Psychophys. **50**, 205–214.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003**). ''One factor underlies individual differences in auditory informational masking within and across age groups,'' Percept. Psychophys. **65**, 396–406.

Micheyl, C., Arthaud, P., Reinhart, C., and Collet, L. (**2000**). ''Informational masking in normal-hearing and hearing-impaired listeners,'' Acta Oto-Laryngol. **120**, 242–246.

Neff, D. L., and Callaghan, B. P. (**1988**). ''Effective properties of multicomponent simultaneous maskers under conditions of uncertainty,'' J. Acoust. Soc. Am. **83**, 1833–1838.

Neff, D. L., and Dethlefs, T. M. (**1995**). ''Individual differences in simultaneous masking with random-frequency, multicomponent maskers,'' J. Acoust. Soc. Am. **98**, 125–134.

Neff, D. L., Dethlefs, T. M., and Jesteadt, W. (**1993**). ''Informational masking for multicomponent maskers with spectral gaps,'' J. Acoust. Soc. Am. **94**, 3112–3126.

Neff, D. L., Kessler, C. J., and Dethlefs, T. M. (**1996**). ''Sex differences in simultaneous masking with random-frequency maskers,'' J. Acoust. Soc. Am. **100**, 2547–2550.

Patterson, R. D. (**1976**). ''Auditory filter shapes derived with noise stimuli,'' J. Acoust. Soc. Am. **59**, 640–654.

Patterson, R. D., and Moore, B. C. J., (**1986**). ''Auditory filters and excitation patterns as representations of frequency resolution,'' in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).

Pollack, I. (**1975**). ''Auditory informational masking,'' J. Acoust. Soc. Am. Suppl. 1 **57**, S5.

Scharf, B., Quigley, S., Aoki, C., Peachy, N., and Reeves, A. (**1987**). ''Focused auditory attention and frequency selectivity,'' Percept. Psychophys. **42**, 215–223.

Soderquist, D. R. (**1970**). ''Frequency analysis and the critical band,'' Psychonomic Sci. **21**, 117–119.

Spiegel, M. F., and Watson, C. S. (**1984**). ''Performance on frequency-discrimination tasks by musicians and nonmusicians,'' J. Acoust. Soc. Am. **76**, 1690–1695.

Stone, M. A., Glasberg, B. R., and Moore, B. C. J. (**1992**). ''Simplified measurement of impaired auditory filter shapes using the notched-noise method,'' Br. J. Audiol. **26**, 329–334.

Watson, C. S., (**1987**). ''Uncertainty, informational masking, and the capacity of immediate auditory memory,'' in *Auditory processing of complex sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, N.J.).

Wright, B. A., and Dai, H. (**1994**). ''Detection of unexpected tones with short and long durations,'' J. Acoust. Soc. Am. **95**, 931–938.

Time course of loudness recalibration: Implications for loudness enhancement

Yoav Arieh[a] and Lawrence E. Marks[b]
John B. Pierce Laboratory and Yale University, New Haven, Connecticut 06519

(Received 20 December 2002; revised 25 June 2003; accepted 7 July 2003)

Loudness recalibration, the effect of a relatively loud 2500-Hz recalibrating tone on the loudness of a relatively soft 2500-Hz target tone, was measured as a function of the interstimulus interval (ISI) between them. The loudness of the target tone, assessed by a 500-Hz comparison tone, declined when the ISI equaled or exceeded about 200 ms and leveled off at an ISI of about 700 ms. Notably, the target tone's loudness did not change significantly at very short ISIs (<150 ms). The latter result is incompatible with the literature reporting loudness enhancement in this time window, but is compatible with the suggestion made by Scharf, Buus, and Nieder [J. Acoust. Soc. Am. **112**, 807–810 (2002)] that early measurements of enhancement were contaminated by the influence of the recalibrating tone on the comparison tone when the two shared the same frequency. In a second experiment the frequency of the comparison tone was changed to 2500 Hz and the results of a loudness enhancement paradigm was successfully predicted from the time course of recalibration obtained in experiment 1. [DOI: 10.1121/1.1603768]

PACS numbers: 43.66.Mk [NFV]

I. INTRODUCTION

The loudness of an auditory signal is intimately tied to the presence of other auditory signals recently presented. In loudness recalibration, listeners experience a reduction in the loudness of transient tones of relatively moderate intensity when stronger tones were recently presented (Marks, 1988, 1994). For example, Marks (1993) exposed listeners to a series of repeated tones: 500 Hz at 53 dB, 500 Hz at 73 dB, 2500 Hz at 48 dB, or 2500 Hz at 68 dB. After exposure, listeners compared previously matched test tones at 500 and 2500 Hz. Exposure to only the greater SPL at each frequency influenced the subsequent judgments. Exposure to the 500-Hz tone at 73 dB decreased the probability of judging a subsequent 500-Hz tone as louder than a 2500-Hz tone, and exposure to the 2500-Hz tone at 68 dB increased the probability. Exposure to the softer tones had essentially no effect.

Although loudness recalibration was first reported 15 years ago, not much is known about the underlying process—perhaps in part because early interpretations of loudness recalibration placed its origin in high-level cognitive processes, as numerical response bias in magnitude estimation (e.g., Marks, 1988) or as shifts in response criteria in loudness matching (see Arieh and Marks, 2001). Subsequent research, however, has shown both explanations to be inadequate. Recalibration arises in paradigms that do not ask listeners to make numerical judgments, either by having participants compare differences in loudness (Schneider and Parker, 1990) or match loudness (Marks, 1992, 1993, 1994; Mapes-Riordan and Yost, 1999; see Scharf *et al.* 2002). Both comparisons of loudness difference and direct loudness comparisons reveal recalibration, much like that observed with magnitude estimation. Finally, in this regard, Arieh and Marks (2003) found recalibration in measures of response times and errors obtained in a speeded choice task. When listeners rapidly classified 500- and 2500-Hz tones as low or high in frequency while the tones took on different SPLs in different conditions, choice responses were longer and errors generally greater under conditions in which loudness was smaller (recalibrated). The positive correlation between choice response time and error rate is the hallmark of a shift in ''sensitivity'' rather than a change in criterion. Thus, loudness recalibration is best conceptualized as a sensory phenomenon that influences auditory responsiveness, reducing loudness of suprathreshold tones (we note here that the term Induced Loudness Reduction has also been used to describe the phenomenon at hand, see Scharf *et al.*, 2002).

Relatively little is known, however, about several basic properties of loudness recalibration, especially its temporal properties: How long does it take for recalibration to arise? How long does it last? The current study aims to elucidate the temporal relationship between the first signal in each sequence, the relatively strong recalibrating tone, and the second signal, the weaker target tone. Here we ask: What is the shortest time interval between recalibrating and target tones that produces loudness recalibration? What time interval produces the greatest recalibration? [For a preliminary stab at some of these questions, see Mapes-Riordan and Yost (1998).]

To evaluate the time course of recalibration, we modified the paradigm used by Mapes-Riordan and Yost (1999). Figure 1 illustrates the events constituting a baseline trial and an experimental trial. Each experimental trial consists of the following sequence: recalibrating tone, interstimulus interval 1 (ISI), target tone, interstimulus interval 2, and comparison tone. The comparison tone is the yardstick for measuring recalibration of the target tone. In baseline trials, the recalibrating tone is omitted, so recalibration is computed as the difference in dB between the levels of the comparison tone

[a]Address for correspondence: The John B. Pierce Laboratory, 290 Congress Avenue, New Haven, CT 06519. Electronic mail: yarieh@jbpierce.org
[b]Electronic mail: marks@jbpierce.org

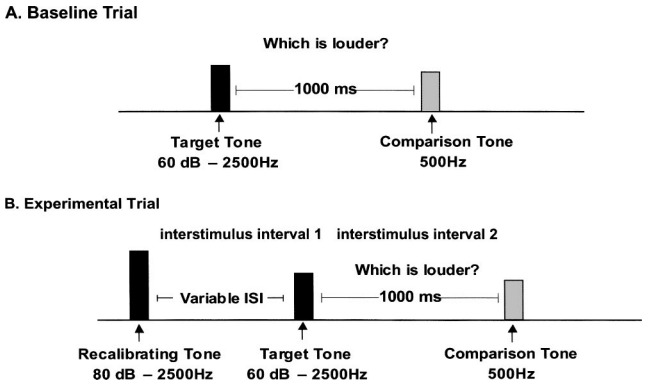


FIG. 1. Schematic illustration of stimulus sequences used to measure loudness recalibration. The upper sequence (a) shows a baseline trial, where a match was determined between the target tone and the comparison tone, and the lower sequence (b) shows an experimental trial, where the recalibrating tone precedes the target tone.

that match the loudness of the target in experimental and baseline trials.

In experimental trials, the listener's task is to ignore the recalibrating tone and judge whether the target tone or the comparison tone is louder. Note that the frequency of the recalibrating tone and the target tone is the same, 2500 Hz, while the frequency of the comparison tone is markedly different, 500 Hz. Although heterofrequency comparison of loudness is difficult, it is necessary in order to minimize the effect of the recalibration tone on the comparison tone. Marks and Warner (1991) and Marks (1994) showed that recalibration essentially affects all signals falling within roughly a critical bandwidth. The difference between the frequency of the recalibrating tone (and test tone) and the frequency of the comparison tone must be substantial, lest the recalibrating tone affect the loudness of the comparison.

In experiment 1 listeners compared the loudness of the target to the loudness of the comparison in a randomized adaptive two-track procedure. In the baseline session, we measured the matching point between the 2500-Hz target at 60 dB and the 500-Hz comparison prior to the induction of recalibration. In the experimental sessions, a 2500-Hz recalibrating tone at 80 dB preceded each comparison trial. Ten ISIs were tested, each in a different session, the shortest being 50 ms and the longest being 3300 ms.

II. EXPERIMENT 1—TIME COURSE OF LOUDNESS RECALIBRATION

A. Method

Ten listeners, six women and four men, participated in experiment 1. All were Yale undergraduates or employees of the J. B. Pierce Laboratory, 19 to 35 years of age, who reported normal hearing.

Participants sat in a sound treated booth. A Tucker-Davis System 3 Real Time processor at a sampling frequency of 50 kHz, driven by a Matlab program running on a Pentium III PC, produced the stimuli in both experiments 1 and 2. The 2500-Hz recalibration and target tones and the 500-Hz comparison tone, appropriately attenuated (Tucker-Davis PA5 module) and gated (5 msec cosine^2 rise and decay), were delivered binaurally for 50 ms through calibrated TDH-49 headphones mounted in MX41/AR cushions. The Matlab program also recorded the listeners' responses and provided all other aspects of user interface for the experimental session.

Before the start of the experiment, each listener received written and oral instructions. The baseline session always came first. A randomized adaptive two-track ascending and descending procedure [two-up, two-down, a variant of the procedure described in Jesteadt (1980)] served to estimate the value of the 500-Hz tone equal in loudness to the 2500-Hz target at 60 dB. Each baseline trial presented a fixed level of the target followed after 1 s by a comparison tone [see Fig. 1(a)]. The listener's task was to judge which tone was louder by pressing an appropriate key. The level of the comparison tone was contingent on the listener's response. If the listener indicated on two successive trials that the target was louder than the comparison, then on the next trial the level of the comparison increased. Alternatively, if the listener indicated twice in succession that the comparison was louder than the target, then on the next trial the level of the comparison decreased. The size of the step started at 4 dB, then decreased to 2 dB after three reversals of direction. In each baseline session, two tracks were randomly interleaved; an ascending track began with the comparison set below the target, at 40 dB, and the descending set above the target, at 80 dB.

The matching point between the target and comparison was calculated by averaging the last six reversal points out of the nine recorded for each track. Thus, for each listener the matching point was based on the average of 12 SPLs (6 per track). The baseline session lasted approximately 10 min.

The procedure for the ten experimental sessions was identical to that of the baseline session except that a 2500-Hz recalibration tone at 80 dB SPL preceded each trial [see Fig. 1(b)]. The ISI between the recalibrating tone and the target tone varied over the sessions: 50, 75, 150, 225, 375, 525, 675, 825, 1650, and 3300 ms. The listeners were instructed to ignore the recalibrating tone and judge which tone was louder, the target or the comparison. For each ISI, recalibration was computed as the difference between the matching levels obtained with and without the recalibrating tone.

In the experimental session, the starting points for the ascending and descending tracks were set individually 25 dB below and above that listener's baseline. Listeners took part in the 11 sessions (baseline plus ten ISIs) over three different days, with no more than four sessions on a single day. Each session lasted 10–15 min, and a mandatory 15-min break separated consecutive sessions. After the baseline session, the ten experimental sessions were given in a different random order for each listener.

B. Results

Figure 2 presents the individual data of the ten listeners who participated in experiment 1. Each panel shows how the matching SPL varied with ISI; in each case, the solid line shows the individual baseline.

The average baseline was 61.1 dB (SE=1.55). The slightly greater SPL at 500 Hz vs 2500 Hz is consistent with

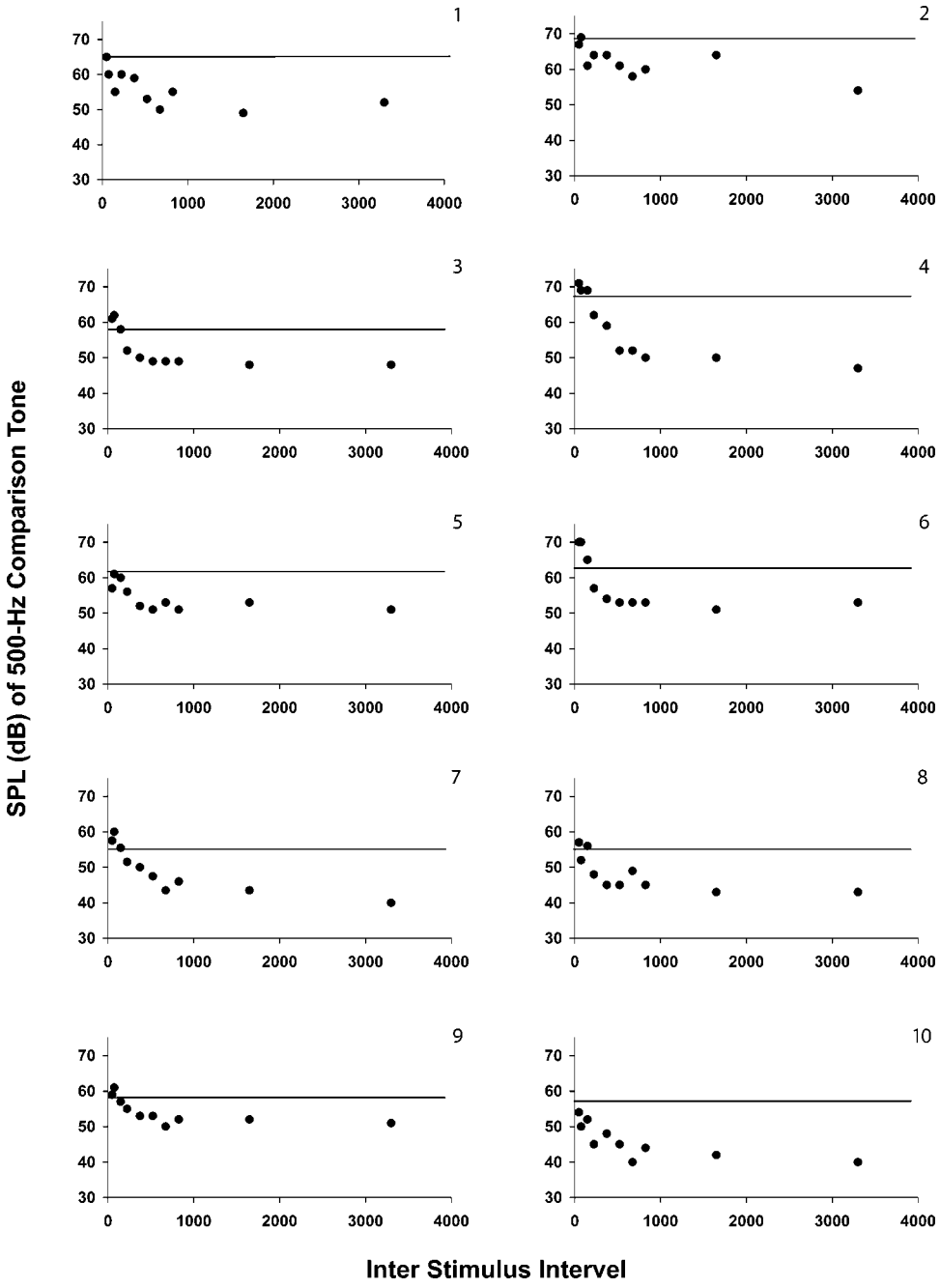


FIG. 2. Individual results from the ten listeners who took part in experiment 1. Plotted for each listener is the matching level in dB obtained for each of ten ISIs. The solid line represents, for each listener, the baseline level obtained in the absence of the recalibrating tone.

equal-loudness functions (e.g., Fletcher and Munson, 1933). In general, there is little or no evidence of recalibration (loudness below baseline) at the smallest ISIs. Recalibration becomes clear only with ISIs greater than 150 ms. The overall pattern is clarified in Fig. 3, which shows results averaged across the ten listeners. A 95% confidence interval is plotted around each mean. These within-participant confidence intervals were derived from the interaction error term of participant by conditions, as suggested by Loftus and Masson (1994). The horizontal solid line shows the mean baseline matching value and the dashed lines indicate 95% confidence intervals around that mean. Values within brackets show loudness recalibration, in dB, at each ISI.

At the three shortest ISIs (50, 75, and 150 ms), recalibration is negligible as the mean matching levels fall within or slightly below the confidence intervals at baseline. Recalibration first becomes significant, and equal to 6.1 dB, at an ISI of 225 ms. Recalibration then increases steadily until it reaches a more or less constant value of 11 dB at an ISI of 675 ms. Recalibration increases slightly again, to 13 dB, at an ISI of 3300 ms, but this increase may be spurious. Overall, this picture—the first detailed account of the time course of recalibration—reveals a temporally graded process that appears to start about 200 ms after the recalibrating tone. Recalibration then increases monotonically in magnitude for about 500 ms and then levels off for at least 3 s or so.

Of special interest is the lack of substantial change in loudness over the three shortest ISIs. Although the present results are consistent with the existence of a modest, continuous change in loudness starting with an ISI as small as 50

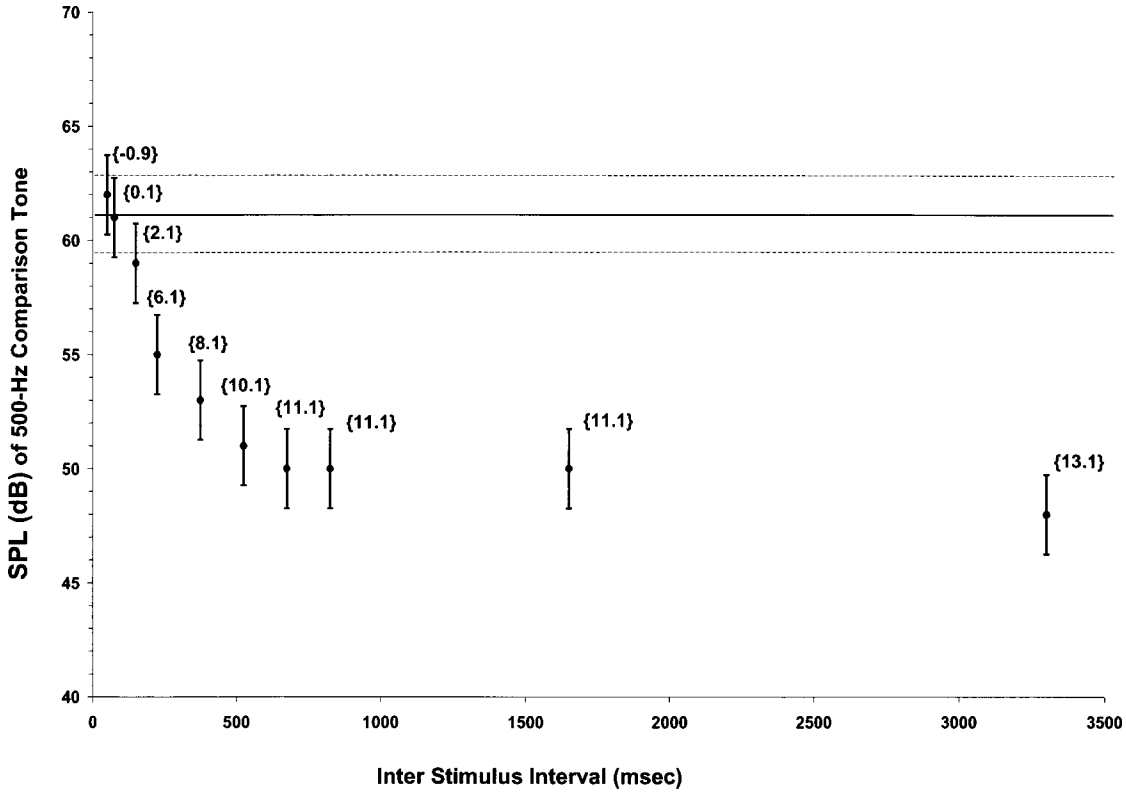


FIG. 3. Pooled matching points (in dB) for the ten listeners in experiment 1, plotted against their respective ISI. Error bars around each point indicate 95% confidence intervals. The solid line represents the mean baseline level obtained in the absence of the recalibrating tone and the dotted lines indicate its 95% confidence intervals.

ms, changes over the first 200 ms or so are small, amounting to no more than 2–3 dB. This outcome is surprising because it is at odds with reports of loudness enhancement. Several studies have shown that when two tones are presented in close temporal proximity—with ISIs<100 ms—the loudness of the second tone appeared to be elevated, if the first tone is more intense than the second (Elmasian and Galambos, 1975; Plack, 1996; Zeng, 1994; Zwislocki and Sokolich, 1974). Most studies of enhancement used a three-tone paradigm like that used in experiment 1. In those studies, the first tone purportedly enhances the loudness of the second, target, tone, measured by comparing its loudness to that of a third, comparison tone. Elmasian and Galambos (1975) and Zwislocki and Sokolich (1974) had listeners adjust the level of the third tone such to equal the loudness of the second tone. Plack (1996) and Zeng (1994) asked listeners to judge which tone was louder, the second or the third, in a two-alternative forced choice adaptive procedure much like ours. In both cases the outcome was the same. The loudness of the second tone was elevated in the presence of the stronger first tone. The magnitude of loudness enhancement has been reported to reach 14 to 16 dB under optimal conditions (Zeng, 1994), but is usually about 10 dB. Why the palpable discrepancy between the substantial changes in loudness at short ISIs reported in studies of loudness enhancement and the lack of any substantial change at comparable ISIs here?

The answer may lie in a suggestion recently made by Scharf *et al.* (2002). Surveying the literature on loudness enhancement, Scharf *et al*. note that virtually all studies used a comparison tone with the same frequency as the two experimental tones. By doing so, Scharf *et al.* suggest, it is possible that the loudness of the comparison tone is affected (recalibrated) by the inducing tone. For example, Zeng (1994) used 1000-Hz signals in his three-tone paradigm, the interval between the first (recalibrating) tone and the third (comparison) tone being 750 ms. When the first tone was 20 dB greater than the second (target), Zeng (1994) found loudness enhancement of 9 dB. According to Fig. 3, at an ISI of 750 ms, the loudness of the third tone would be reduced by 11 dB. In one condition of Zwislocki and Sokolich's study (1974), the interval between the first and the third tones was 575 ms (75 ms between the first tone and the target and 500 ms between the target and the comparison), and the associated loudness enhancement was reported to be about 10 dB. According to Fig. 3, however, the loudness of the comparison tone should have reduced by this same amount. In the study of Zwislocki and Sokolich, the magnitude of the loudness enhancement dropped to zero when the interval between the first tone and the second was 500 ms and so the interval between the first tone and the third tone was 1000 ms. According to Fig. 3, at these ISIs, both the target and the comparison should have been recalibrated by about 10 dB, making them equal in loudness.

This little analysis supports the suggestions made by Scharf *et al.* (2002), namely that so-called loudness enhancement is the outcome of a reduction (recalibration) in the loudness of the comparison tone. A strong first tone can reduce the loudness of the comparison, making it necessary to increase its level to make it as loud as the target. The target itself is little affected when the ISI between the initial reca-

TABLE I. Predicted and observed loudness enhancement for the four listeners in experiment 2, based on the time course of their loudness recalibration in experiment 1. The 95% confidence intervals are in parentheses.

ISI (ms)	Predicted loudness recalibration for the target tone (dB)	Predicted loudness recalibration for the comparison tone (dB)	Predicted loudness enhancement (dB)	Observed loudness enhancement (dB)
75	2	11.25	9.25(1.7)	11(1.7)
525	10	13.20	3.20(1.5)	2(1.5)

librating tone and the target is small. In their report, Scharf *et al.* (2002) presented preliminary data that tentatively supported their argument. In experiment 2 of this study, we took a more thorough look at this account of loudness enhancement, bringing back four of the listeners who participated in experiment 1 to test more explicitly the idea that the magnitude of loudness enhancement can be predicted from the time course of loudness recalibration.

III. EXPERIMENT 2—PREDICTING LOUDNESS ENHANCEMENT FROM THE TIME COURSE OF LOUDNESS RECALIBRATION

A. Method

Four listeners, two women and two men, participated in experiment 2. All four had taken part in experiment 1. The experimental apparatus and experimental paradigm were identical to those used in experiment 1 except for one crucial detail—in experiment 2, the frequency of the comparison was set to equal the frequency of the recalibrating and target tones, that is, to 2500 Hz. Only two ISIs were used: 75 and 525 ms. The order of the sessions was counter-balanced over listeners.

B. Results

We are in the unique position of being able to predict quantitatively the amount of enhancement in experiment 2 from the measurements of recalibration obtained in experiment 1. These predictions can then be compared to the actual loudness changes obtained in experiment 2. To predict loudness enhancement for each listener, we first read off, from Fig. 2, the amount of recalibration at the two ISIs used in experiment 2 (75 and 525 ms). These values represent the predicted changes in loudness of the target tone in experiment 2. Second, in the same manner we read off, from Fig. 2, the amount of recalibration predicted for the comparison tone. The comparison tone is presented 1075 ms after the recalibrating tone in the short ISI condition (75 ms between the recalibration tone and the target+1000 ms between the target and the comparison) and 1525 ms after the recalibrating tone in the long ISI condition. Finally, if recalibration is the only process that influences loudness, then the observed ''enhancement'' in experiment 2 should equal the difference between the predicted values of the target and the comparison tone.

Columns 2 and 3 of Table I give the average predicted values of loudness recalibration for the target tone and for the comparison tone, respectively, computed for the four listeners who served in experiment 2. Column 4 gives the predicted amount of loudness ''enhancement,'' which equals the difference between the first two values. This is essentially the difference between the small loudness recalibration of the target and the larger recalibration of the comparison. Finally, the last column gives the average observed ''enhancement.''

Two features of the results are salient. First, loudness enhancement was substantially greater at an ISI of 75 ms than at 525 ms. This outcome is compatible with previous reports (Elmasian and Galambos, 1975; Zwislocki and Sokolich, 1974). Second, the measures of loudness enhancement fall very close to the predictions from experiment 1. This outcome confirms the suggestion made by Scharf *et al.* (2002) that the reported loudness enhancement of a weaker tone by a stronger one at short ISIs is the result of a reduction in loudness (recalibration) of the comparison tone. The loudness of the comparison is affected by the stronger recalibration tone when their frequencies are the same.

IV. GENERAL DISCUSSION

The present study provides two major new findings: (a) the first systematic description of the onset and time course of loudness recalibration, and (b) an empirical confirmation of the hypothesis put forward by Scharf *et al.* (2002) that loudness enhancement as measured with a three-tone single frequency paradigm reflects a reduction in the loudness of the comparison rather than an increase in the loudness of the target.

A significant amount of loudness recalibration first appeared about 200 ms after the presentation of the recalibration tone. The magnitude of loudness reduction then increased monotonically with ISI until approximately 675 ms, after which the loudness reduction leveled off at a magnitude of 11 dB. This magnitude is compatible with previous studies (Marks, 1988, 1994) that reported loudness recalibration between 17 and 22 dB when computed as the overall relative shift in loudness at two frequencies—500 and 2500 Hz. Assuming that the amount of loudness recalibration is about equal at the two frequencies, we expect to measure half as much recalibration at just one frequency. The present results do not speak to the issue of recovery from recalibration because, at the longest ISI, 3300 ms, recalibration was still substantial. Preliminary data from our lab indicate, however, that recovery times may extend over dozens of seconds and may depend strongly on the number of recalibration tones previously presented.

A. Physiological considerations

The temporal growth of loudness recalibration can serve as an important clue for identifying the underlying physiological process. Generally, the neural response to a test tone is attenuated when a louder tone precedes the test tone (but see Brosch and Schreiner, 2000). This effect is sometimes

called short-term adaptation and is studied using a forward masking paradigm (Boettcher *et al.*, 1990; Shore, 1995). Numerous studies have shown forward inhibition of neural responses at several levels of the auditory system, including the auditory nerve (Smith, 1977; Harris and Dallos, 1979), cochlear nuclear complex (Boettcher *et al.*, 1990), inferior colliculus (Aitkin and Dunlop, 1969; Etholm, 1969), medial geniculate body (Aitkin and Dunlop, 1969; Schreiner, 1981), and auditory cortex (Brosch and Schreiner, 1997; Hocherman and Gilat, 1981). Recovery times of the neural response, however, are shorter by an order of magnitude than the time course of loudness recalibration found here. Recovery times average about 100 ms at the auditory nerve (Harris and Dallos, 1979; Smith, 1977), slightly longer at the cochlear nucleus (Shore, 1995) and even longer, around 200–250 ms, at the medial geniculate (Schreiner, 1981). The longest neural recovery times, recorded in the auditory cortex, extend up to only 1600 ms (Hocherman and Gilat, 1981).

These findings, however, should be interpreted cautiously. First, neural measures have been made in different species and under various kinds of anesthesia, sometimes under no anesthesia at all. Second, stimuli used in different studies varied considerably in their duration, intensity, and frequency. So direct comparison of neural recovery times across studies requires caution. This said, the general trend indicates longer recovery times as we travel from the periphery to the central auditory system. Thus, although none of the reported recovery times matches that of recalibration, the higher, central levels of the auditory system are the most promising candidates for providing neural correlates in time constants. This view is compatible with the view that recovery times are greater at higher levels of the auditory system because inhibitory processes tend to accumulate as auditory stimulation travels from the periphery to the central nervous system. This happens partly because central auditory regions show increasing numbers of local circuits and interconnecting neurons that can locally inhibit sensitivity to tones (Brosch and Schreiner, 1997). Thus, in sum, central inhibition provides one possible mechanism of recalibration.

Parker and Schneider (1994) have suggested another possible mechanism of loudness recalibration, involving descending efferent pathways. They suggest that a nonlinear amplifier, controlled by a top-down mechanism, attenuates incoming signals whenever they include loud tones. One example of descending auditory pathway is the efferent system that projects from the medial olivocochlear neurons to the outer hair cells (Liberman, 1986). Feedback through the cochlear efferent system reduces vibration at the basilar membrane (Russell and Murugasu, 1997) and activity in the auditory nerve (Guinan and Stankovic, 1996) in response to stimulation. Two features of the efferent system make it an attractive candidate for explaining loudness recalibration (Nieder *et al.*, 2003). First, the efferent system responds more vigorously to ipsilateral than contralateral stimulation (Liberman, 1988), consistent with evidence that recalibration is greater when induced ipsilaterally rather than contralaterally (Marks, 1996). Second, the efferent system can be sensitized for a few minutes by exposure to relatively intense broadband noise (Liberman, 1988). This sensitization might serve to reduce afferent responses to acoustic stimuli. But an explanation in terms of efferent activity also has some weaknesses. Because the efferent system exerts it influence early in auditory processing, one would expect to find evidence of attenuation (recalibration) at distal auditory loci. The available evidence, summarized above, fails to show the long recovery times in the auditory nerve that one would expect if recalibration resulted from efferent feedback to the cochlea. Thus, even if the efferent system contributes significantly to the loss of sensitivity following stimulation, it cannot be the sole source of loudness recalibration.

B. Loudness recalibration and loudness enhancement

The results of experiment 1 show that loudness recalibration is durable, extending more than 3 s after the presentation of a brief recalibrating tone. Further, as already mentioned, recalibration is frequency specific. The combination of durability and frequency specificity constrains the methods that can be used to measure loudness recalibration through matching procedures. Notably, one cannot use a comparison tone with the same frequency as the recalibrating tone and the target tone. If one does, the recalibrating tone will affect the loudness of the comparison tone, thereby distorting the measurements. According to Scharf *et al.* (2002), this is exactly what has happened in studies of loudness enhancement. Those studies reported that the loudness of the target tone is enhanced when presented up to 100 ms after a more intense recalibrating tone (or the conditioning tone, as it was termed). However, the frequency of the comparison was always the same as that of the recalibrating tone, which it usually followed by less than 2 s. Scharf *et al.* (2002) suggested that (a) the loudness of the comparison tone was reduced by the recalibrating tone and, therefore, (b) listeners had to raise the intensity of comparison in order to match its loudness to that of the target. Thus, loudness enhancement was an artifact of the way it was measured. In experiment 1, we used a comparison tone that lies almost seven critical bands from the recalibrating and target tones. The absence of substantial loudness enhancement at short ISIs supports Scharf *et al.*'s position.

Using loudness reduction to account for traditional measures of loudness enhancement bears consequences for other findings that have been linked to it. For example, loudness discrimination of two test tones following a ''priming'' tone is impaired (jnd's increase) when the level of the priming tone exceeds that of the first test tone by about 20–30 dB but not when the difference is larger or smaller (Plack, 1996; Zeng, 1994). Perceptual variability associated with the presumed enhancing effect of the primer on the first target was suggested as a possible source for the elevation in jnd's. Instead of enhancement, we would argue that the priming (recalibrating) tone differentially reduces the loudness of the two test tones, the second more than the first.

Note that loudness enhancement has also been reported in vibrotactile sensation. When two vibrations are presented successively to the same site, the subjective magnitude of the second is enhanced if the ISI between the two is about 100 ms (Gescheider and Verrillo, 1982; Verrillo and Gescheider, 1975, 1976). Vibrotactile enhancement resembles auditory

loudness enhancement in both time course and magnitude. Unfortunately, studies of vibratory enhancement have had the same methodological flaw as studies of auditory enhancement, namely, the use of a three-stimulus paradigm in which the comparison tone had the same frequency as the recalibrating and target tones. We propose here that, as in hearing, vibrotactile loudness enhancement will disappear if measured by a comparison tone whose frequency differs significantly from that of the recalibrating and target tones.

Does all of this mean that loudness enhancement does not exist? Results of experiment 1 show that at short ISIs, any change in the loudness of the target is small. In other words, the recalibration tone substantially affects the loudness of a tone presented 200 ms later, but affects only slightly if at all the loudness of a tone presented 75 ms later. One possible explanation to this puzzle is that it takes about 200 ms for the inhibitory process responsible for recalibration to develop. But there is another possibility, one that assumes no temporal threshold for the onset of recalibration. The lack of a reliable change in loudness at short ISIs could reflect two offsetting effects: inhibition (recalibration) and a small enhancement (with the underlying processes having different time constants). This possibility is especially cogent in light of the physiological evidence showing inhibitory neural processes at all levels of the auditory system at ISIs shorter than 100 ms. The absence of a suprathreshold perceptual analog to the neural inhibition observed at short ISIs could mean that a short-lived facilitory process of loudness enhancement offsets an inhibitory process of loudness recalibration.

ACKNOWLEDGMENTS

This research was supported by NIH Grant No. DC 03842-04 from the National Institutes of Deafness and Other Communication Disorders to LEM. We thank B. Scharf, Neal Viemeister, and an anonymous reviewer for cogent and thoughtful comments. Yoav Arieh is now at the Psychology department, Montclair State University, Upper Montclair, New Jersey.

Aitkin, L. M., and Dunlop, C. W. (**1969**). ''Inhibition in the medial geniculate body of the cat,'' Exp. Brain Res. **7**, 68–83.

Arieh, Y., and Marks, L. E. (**2001**). ''Recalibration of loudness: sensory vs. decisional processes,'' in *Fechner Day 2001. Proceedings of the Seventeenth Annual Meeting of the International Society for Psychophysics*, edited by T. Sommerfeld, R. Kompass, and T. Lachman (Pabst Science, Berlin), pp. 242–247.

Arieh, Y., and Marks, L. E. (**2003**). ''Recalibrating the auditory system: a speed-accuracy analysis of intensity perception,'' J. Exp. Psychol. Human Percept. Perform. **29**, 523–536.

Boettcher, F. A., Salvi, R. J., and Saunders, S. S. (**1990**). ''Recovery from short-term adaptation in single neurons in the cochlear nucleus,'' Hear. Res. **48**, 125–144.

Brosch, M., and Schreiner, C. E. (**1997**). ''Time course of forward masking tuning curves in cat primary auditory cortex,'' J. Neurophysiol. **77**, 923–943.

Brosch, M., and Schreiner, C. E. (**2000**). ''Sequence sensitivity of neurons in cat primary auditory cortex,'' Cereb. Cortex **10**, 1155–1167.

Elmasian, R., and Galambos, R. (**1975**). ''Loudness enhancement: monaural, binaural, and dichotic,'' J. Acoust. Soc. Am. **58**, 229–234.

Etholm, B. (**1969**). ''Evoked responses in the inferior colliculus, medial geniculate body and auditory cortex by single and double clicks in cats,'' Acta Oto-Laryngol. **67**, 319–325.

Fletcher, H., and Munson, W. A. (**1933**). ''Loudness, its definition, measurement and calculation,'' J. Acoust. Soc. Am. **5**, 82–108.

Gescheider, G. A., and Verrillo, R. T. (**1982**). ''Contralateral enhancement and suppression of vibrotactile sensation,'' Percept. Psychophys. **32**, 69–74.

Guinan, J. J., and Stankovic, K. M. (**1996**). ''Medial efferent inhibition produces the largest equivalent attenuations at moderate to high sound levels in cat auditory-nerve fibers,'' J. Acoust. Soc. Am. **100**, 1680–1690.

Harris, D. M., and Dallos, P. (**1979**). ''Forward masking of auditory nerve fiber responses,'' J. Neurophysiol. **42**, 1083–1107.

Hocherman, S., and Gilat, E. (**1981**). ''Dependence of auditory evoked unit activity on interstimulus interval in the cat,'' J. Neurophysiol. **45**, 987–997.

Jesteadt, W. (**1980**). ''An adaptive procedure for subjective judgments,'' Percept. Psychophys. **28**, 85–88.

Liberman, M. C. (**1988**). ''Response properties of cochlear efferent neurons: monaural vs. binaural stimulation and the effects of noise,'' J. Neurophysiol. **60**, 1779–1798.

Liberman, M. C., and Brown, M. C. (**1986**). ''Physiology and anatomy of single olivocochlear neurons in the cat,'' Hear. Res. **24**, 17–36.

Loftus, G. R., and Masson, M. E. J. (**1994**). ''Using confidence intervals in within-subjects designs,'' Psychon. Bull. Rev. **1**, 476–490.

Mapes-Riordan, D., and Yost, W. A. (**1998**). ''Temporal properties of loudness recalibration,'' J. Acoust. Soc. Am. **103**, 3020.

Mapes-Riordan, D., and Yost, W. A. (**1999**). ''Loudness recalibration as a function of level,'' J. Acoust. Soc. Am. **106**, 3506–3511.

Marks, L. E. (**1988**). ''Magnitude estimation and sensory matching,'' Percept. Psychophys. **43**, 511–525.

Marks, L. E. (**1992**). ''The slippery context effects in psychophysics: intensive, extensive, and qualitative continua,'' Percept. Psychophys. **51**, 187–198.

Marks, L. E. (**1993**). ''Contextual processing of multidimensional and unidimensional auditory stimuli,'' J. Exp. Psychol. Hum. Percept. Perform. **19**, 227–249.

Marks, L. E. (**1994**). ''Recalibrating' the auditory system: the perception of loudness,'' J. Exp. Psychol. Hum. Percept. Perform. **20**, 382–396.

Marks, L. E. (**1996**). ''Recalibrating the perception of loudness: interaural transfer,'' J. Acoust. Soc. Am. **100**, 473–480.

Marks, L. E., and Warner, E. (**1991**). ''Slippery context effect and critical bands,'' J. Exp. Psychol. Hum. Percept. Perform. **17**, 986–996.

Nieder, B., Buus, S., Florentine, M., and Scharf, B. (**2003**). ''How induced loudness reduction depends on duration, level, and laterality,'' J. Acoust. Soc. Am.

Parker, S., and Schneider, B. A. (**1994**). ''The stimulus range effect: evidence for top-down control of sensory intensity in audition,'' Percept. Psychophys. **56**, 1–11.

Plack, C. J. (**1996**). ''Loudness enhancement and intensity discrimination under forward and backward masking,'' J. Acoust. Soc. Am. **100**, 1024–1030.

Russell, I. J., and Murugasu, E. (**1997**). ''Medial efferent inhibition suppresses basilar membrane responses to near characteristic frequency tones of moderate to high intensities,'' J. Acoust. Soc. Am. **102**, 1734–1738.

Scharf, B., Buus, S., and Nieder, B. (**2002**). ''Loudness enhancement: induced loudness reduction in disguise?'' J. Acoust. Soc. Am. **112**, 807–810.

Schneider, B., and Parker, S. (**1990**). ''Does stimulus context affect loudness or only loudness judgment?'' Percept. Psychophys. **48**, 409–418.

Schreiner, C. E. (**1981**). ''Poststimulatory effects in the medial geniculate body of guinea pigs,'' in *Neuronal Mechanisms of Hearing*, edited by J. Syka and L. Aitkin (Plenum, New York), pp. 191–196.

Shore, S. E. (**1995**). ''Recovery of forward masked responses in ventral cochlear nucleus neurons,'' Hear. Res. **82**, 31–43.

Smith, R. L. (**1977**). ''Short-term adaptation in single auditory nerve fibers: Some post-stimulatory effects,'' J. Neurophysiol. **40**, 1098–1112.

Verrillo, R. T., and Gescheider, G. A. (**1975**). ''Enhancement and summation in the perception of two successive vibrotactile stimuli,'' Percept. Psychophys. **18**, 128–136.

Verrillo, R. T., and Gescheider, G. A. (**1976**). ''Effect of double ipsilateral stimulation on vibrotactile sensation magnitude,'' Sens Processes **1**, 127–137.

Zeng, F.-G. (**1994**). ''Loudness growth in forward masking: Relation to intensity discrimination,'' J. Acoust. Soc. Am. **96**, 2127–2132.

Zwislocki, J. J., and Sokolich, W. G. (**1974**). ''On loudness enhancement of a tone burst by a preceding tone burst,'' Percept. Psychophys. **16**, 87–90.

The masking level difference for signals placed in masker envelope minima and maxima

Emily Buss,[a] Joseph W. Hall III, and John H. Grose
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 28 October 2002; revised 9 June 2003; accepted 11 June 2003)

Previous data on the masking level difference (MLD) have suggested that NoSπ detection for a long-duration signal is dominated by signal energy occurring in masker envelope minima. This finding was expanded upon using a brief 500-Hz tonal signal that coincided with either the envelope maximum or minimum of a narrow-band Gaussian noise masker centered at 500 Hz, and data were collected at a range of masker levels. Experiment 1 employed a typical MLD stimulus, consisting of a 30-ms signal and a 50-Hz-wide masker with abrupt spectral edges, and experiment 2 used stimuli generated to eliminate possible spectral cues. Results were quite similar for the two types of stimuli. At the highest masker level the MLD for signals coinciding with masker envelope minima was substantially larger than that for signals coinciding with envelope maxima, a result that was primarily due to decreased NoSπ thresholds in masker minima. For most observers this effect was greatly reduced or eliminated at the lowest masker level. These level effects are broadly consistent with the presence of physiological background noise and with a level-dependent binaural temporal window. Comparison of these results with predictions of a published model suggest that basilar-membrane compression alone does not account for this level effect. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1598199]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Dc [LRB]

I. INTRODUCTION

The masking level difference (MLD) paradigm is one in which the masked detection threshold for a signal is determined as a function of the relative interaural differences between the signal and masker (Hirsh, 1948). In the typical example of the MLD, thresholds for a diotic stimulus, with the signal and masker in phase across the two ears (NoSo), are compared to thresholds obtained when the masker is diotic but the signal is 180° out of phase at the two ears (NoSπ). Under some conditions, thresholds in the NoSπ condition are substantially lower than those in the NoSo condition, particularly at low signal frequencies. The MLD for a 500-Hz signal and broadband masker is commonly reported to be 12 to 15 dB.

Two recent studies (Eddins and Barber, 1998; Hall *et al.*, 1998) examined the effect of masker fluctuation on the magnitude of the MLD. Both studies compared results for a ''low-noise'' masker, a stimulus that is characterized by minimal envelope fluctuation, against those obtained with the more typical Gaussian noise masker. For the NoSo condition, thresholds were lower for a low-noise masker than for a Gaussian masker, consistent with expectations based on previous data collected with monaural stimuli (e.g., Hartmann and Pumplin, 1988). In contrast, in the NoSπ condition thresholds for a low-noise masker were either comparable to, or higher than, those obtained with a Gaussian masker. It was hypothesized in these reports that poorer performance for the NoSπ condition with the low-noise masker is due to the fact that low noise has less pronounced inherent amplitude modulation minima; the epochs of improved signal-to-noise ratio associated with the minima of Gaussian noise provide a detection advantage in the NoSπ condition.

A study by Grose and Hall (1998) estimated temporal weighting functions for a pure-tone signal masked by Gaussian bandpass noise in both the NoSo and NoSπ conditions using a conditional on a single sample (COSS) analysis (Berg, 1989). For the NoSπ condition, that study reported higher perceptual weights for epochs coinciding with masker modulation minima than masker modulation maxima. In contrast, weights for the NoSo condition were approximately equivalent for epochs coinciding with masker modulation minima and maxima. Failure to make use of epochs of improved signal-to-noise ratio with narrow-band monaural stimuli is sometimes discussed in terms of stimulus uncertainty; in the absence of other cues, inherent masker modulation is basically unpredictable and does not support this type of weighting (Buus *et al.*, 1996). Based on their weighting results, Grose and Hall argued that the signal phase disparity of the NoSπ presentation allowed signal energy coinciding with masker modulation minima to be given a larger perceptual weight than signal energy coinciding with modulation maxima. These results provided further evidence that the auditory system utilizes the improved signal-to-noise ratio during masker minima in the NoSπ condition.

It has been suggested that basilar-membrane compression could account for the higher weights given to modulation minima in binaural paradigms (van de Par and Kohlrausch, 1998). The basic idea is that compression acts to reduce the relative power of the modulation maxima, effectively increasing the relative salience of the signal cue during masker modulation minima. This is reflected in higher tem-

[a]Electronic mail: ebuss@med.unc.edu

poral weights for epochs associated with masker modulation minima and, by extension, to the finding of lower MLD for low-noise maskers. Bernstein, van de Par, and Trahiotis (1999) described a model based on interaural correlation that can account for the low-noise MLD data of Eddins and Barber (1998) and Hall *et al.* (1998). In this model a normalized interaural correlation is calculated after a series of processing stages designed to parallel the stimulus transformations that take place at the auditory periphery, including basilar-membrane compression, low-pass filtering, half-wave rectification, and squaring.

The experiments described below measured NoSo and NoSπ thresholds for brief signals coinciding with masker modulation minima and maxima in order to more directly assess the role of masker fluctuation on the MLD. This manipulation was expected to produce larger MLDs for signals coinciding with masker minima than maxima. A range of stimulus levels was utilized to test the hypothesis that the effect of signal placement (coincident with the masker minimum vs maximum) would become less pronounced at lower masker levels, where the basilar-membrane input/output function has been shown to be less compressive (Ruggero *et al.*, 1997). Conditions employing a more typical longer duration signal were also included for comparison. If the MLD for long-duration stimuli is dominated by signal energy coincident with masker minima, then the pattern of thresholds as a function of level obtained with a long-duration masker should more closely resemble that of the brief signal coincident with a modulation minimum than a modulation maximum.

Experiment 1 used bandpass noise maskers generated in the frequency domain by zeroing out components falling outside the bandpass region, a process that creates relatively spectrally sharp noise bands for the stimulus duration used here. Tonal signals were ramped on and off with 15-ms ramps. This stimulus has the advantage that the masker is representative of those used in the MLD literature, but it has the disadvantage that signal-detection cues based on spectral splatter outside of the masker bandwidth may not have been entirely eliminated. Experiment 2 used convolution in the time domain to restrict stimulus bandwidth and to provide signal ramps. This method eliminates spectral cues, but has the disadvantage of lack of comparable published data and of longer duration signal. Results of experiments 1 and 2 were very similar, suggesting that the potential spectral cue in experiment 1 did not play a material role in the results obtained.

II. EXPERIMENT 1

A. Methods

1. Observers

Six adults participated, ranging in age from 19 to 39 years. All had thresholds of 15 dB HL or better at octave frequencies 250 to 8000 Hz. Observers were practiced in psychophysical tasks and were paid for their time.

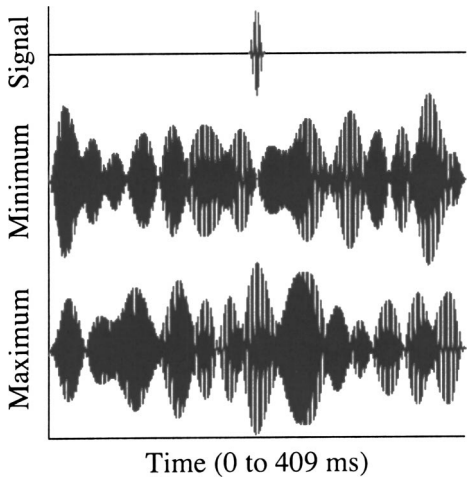


FIG. 1. The top waveform shows a brief signal, such as would be presented in the *min* or *max* conditions. The middle waveform shows a masker token shifted so that the amplitude minimum coincides with the signal, as in the *min* signal condition. The bottom waveform shows the same token shifted so that the amplitude maximum coincides with the signal, as in the *max* signal condition.

2. Stimuli

The signals were 500-Hz tones. In the conditions employing a brief signal coincident with the masker modulation maximum (*max*) or minimum (*min*), the tone was presented in the temporal center of the masker and ramped on and off with 15-ms $\cos^2$ ramps, with no steady state. Example stimuli are shown in Fig. 1. In the *long* condition a 409-ms signal was gated synchronously with the masker. Thresholds are reported in dB SPL of a steady tone of equal peak amplitude, so the brief (*min* and *max* condition) signals and the longer duration signal were not equated for rms level.

Each of the maskers employed in the experiment was generated in the following manner. First, a 409-ms token of 50-Hz-wide Gaussian noise centered on 500 Hz was generated. This token was generated in the frequency domain in a manner that allowed it to be repeated sequentially without abrupt transitions. Next, an extended array was formed consisting of the three replicates of this token, concatenated end to end. The Hilbert envelope of the central 409 ms of the extended array (i.e., the second replicate) was then computed. For the *max* conditions, the maximum value of the envelope was identified, and the extended array was truncated to 409 ms such that this point fell in the temporal center of the truncated array. These manipulations are equivalent to beginning with a single 409-ms array, identifying the maximum value of the envelope, and ''rotating'' the individual digital samples from one end to the other until the desired point is at the temporal center of the token. Similarly, for the *min* conditions the minimum was identified and the extended array was truncated appropriately. Following this manipulation, ramps were applied. The resulting array was 409 ms in duration measured between zero-voltage points. Maskers were generated prior to each interval and presented at 25-, 40-, or 55-dB spectrum level (corresponding to 42, 57, or 72 dB SPL).

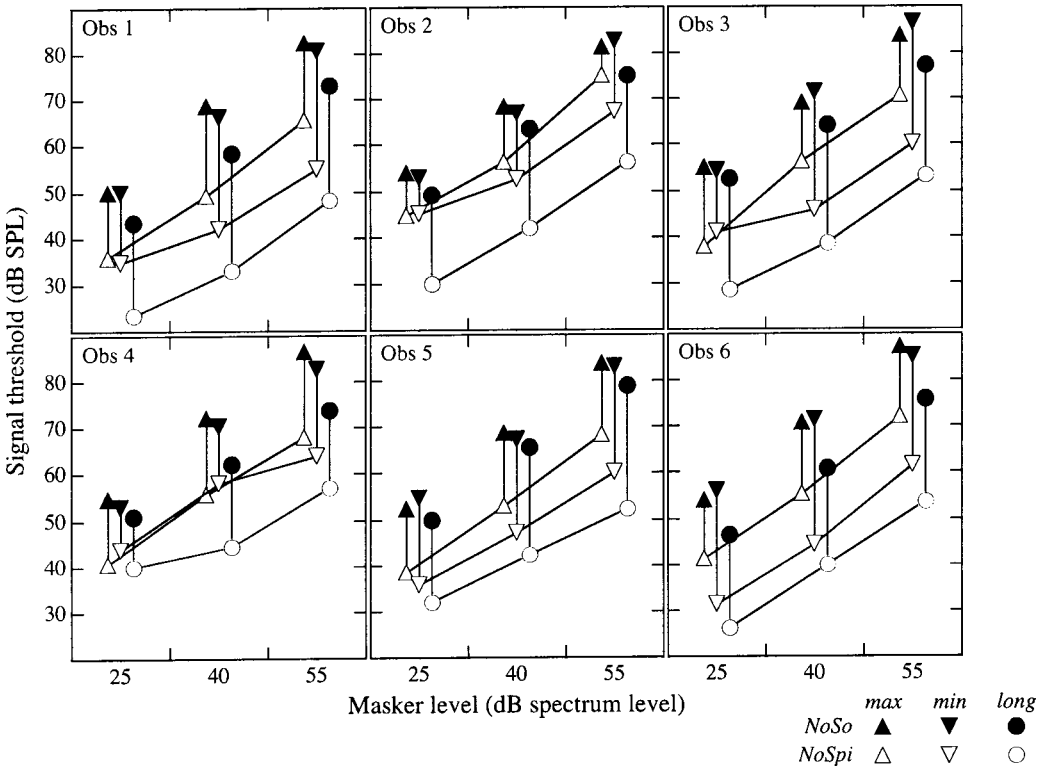


FIG. 2. Mean thresholds are plotted as a function of masker spectrum level. Each panel contains results for one observer, with open symbols indicating NoSπ thresholds for the *max* (△), *min* (▽), and *long* (○) conditions. Filled symbols indicate the associated NoSo thresholds.

3. Procedure

Signal and masker stimuli were played out of two channels of a DAC at 40 kHz, passed through programmable attenuators, summed, and presented over insert phones (Etymotic: ER-2). Signal phase (So vs Sπ) was manipulated by way of an inverter switch on the adder that mixed the signal with the No masker.

Stimuli were presented in a standard 3 alternative-forced-choice, 3-down, 1-up track. Signal level was adjusted in steps of 4 dB for the first two adjustments and then in steps of 2 dB for the remaining six reversals. Thresholds reported here are the average signal level at the last four track reversals. Visual cues to the listening intervals were provided via LEDs which lighted synchronously with masker onset and offset, and as such preceded signal onset by 190 ms for the brief signals. Feedback was given at the end of each trial. Conditions were run in blocks. Three threshold estimates were obtained for each block, with a fourth if the range of estimates exceeded 3 dB. Data reported below are the average of all estimates obtained.

B. Results

Data for each observer are plotted in Fig. 2 as a function of masker level. Symbol shape indicates signal condition, either *max* (△), *min* (▽), or *long* (○). Open symbols indicate thresholds for the NoSπ conditions, filled symbols indicate thresholds for the NoSo conditions, and the length of the vertical connecting lines indicates the magnitude of the MLD. While there are substantial individual differences in the pattern of data, there are some common trends. For the NoSo data, thresholds in the *min* and *max* conditions fell together as a function of masker level, with an average signal-to-noise ratio across observers ranging from 11.3 to 12.3 dB SPL. A repeated-measures ANOVA was performed, with three levels of masker level (25, 40, and 55 dB spectrum level) and two levels of signal condition (*min* and *max*). There was a significant main effect of masker level ($F_{2,10}=2331$, $p<0.0001$), no main effect of signal condition, and no interaction. For the NoSπ data thresholds in the *min* and *max* conditions were more similar at the low masker level, with an advantage in the *min* condition emerging at the higher masker levels. One exception to that trend is the data of Obs 6, who showed a consistent difference between *min* and *max* conditions across the range of masker levels (see Fig. 2). The level effect for a *min* advantage was confirmed with an ANOVA performed on the NoSπ thresholds. This analysis revealed a main effect of masker level ($F_{2,10}=398$, $p<0.0005$), a main effect of signal condition ($F_{1,5}=13.3$, $p<0.05$), and a significant interaction ($F_{2,10}=11.0$, $p<0.005$). These results indicate that the pattern of the MLDs as a function of masker level, discussed below, can be attributed to changes in the NoSπ thresholds rather than a change in the NoSo reference.

Average values of the MLD appear in Fig. 3. The MLD tended to be greatest for the *long* and *min* signal conditions, while the MLD for the *max* condition was reduced. Higher masker level was associated with a larger MLD for the *min* and *long* conditions, but not for the *max* condition. A repeated-measures ANOVA analysis of the MLD data showed a main effect of masker level ($F_{2,10}=15.93$, $p<0.001$) and signal condition ($F_{2,10}=8.70$, $p<0.01$). The interaction between masker level and signal condition was

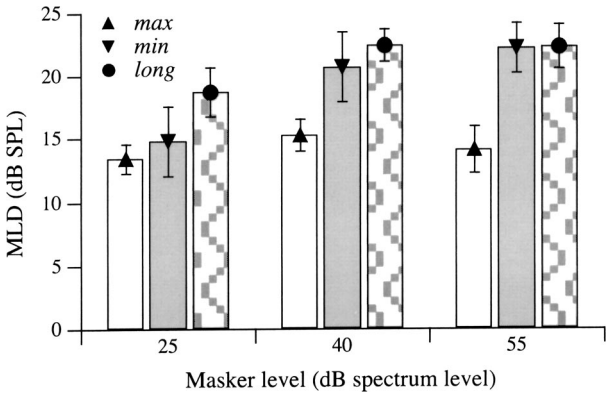


FIG. 3. Mean values of the MLD (dB) are plotted for each signal condition and for each masker level, as indicated along the *x* axis. Symbols indicate the signal condition, as in Fig. 2. Error bars indicate the standard error of the mean ($n=6$).

also significant ($F_{4,20}=2.92, p<0.05$). Simple contrasts showed that MLDs in the *max* and *long* conditions were significantly different ($F_{1,5}=15.87, p<0.05$), but those in the *min* and *long* conditions did not differ reliably from each other ($F_{1,5}=1.71, p=0.25$). This is consistent with the hypothesis that thresholds in the *long* condition are more heavily impacted by signal energy coincident with masker modulation minima, particularly at high masker level.

Consistent with previous data, results here showed that a brief signal presented during a masker minimum was associated with a greater MLD than a signal presented during a masker maximum. This effect was level dependent, being most pronounced for the 55-dB masker level and greatly reduced or absent at the 25-dB masker level for all but one observer. In the literature (Buus *et al.*, 1996), this detection advantage has been discussed in terms of temporal uncertainty as to ''when to listen'' for the signal in the NoSo condition, with the binaural cues eliminating this uncertainty in the NoSπ condition. Before evaluating the level effects in the context of this approach, the possibility of off-frequency listening must be addressed. For these stimuli the 15-ms signal ramps introduce substantial spectral splatter, with signal energy that is only about 3.5 dB down at the edges of the masker passband. Experiment 2 was designed to evaluate whether this spectral cue played a role in the pattern of results observed in experiment 1.

III. EXPERIMENT 2

A. Methods

1. Observers

Four adults participated in experiment 2, ranging in age from 19 to 39 years. All had thresholds 15 dB HL or better at octave frequencies 250 to 8000 Hz. Observers were practiced in psychophysical tasks and were paid for their time. Two observers, Obs 1 and Obs 6, had previously participated in experiment 1.

2. Stimuli

Stimuli were similar to those used in experiment 1, with a narrow-band masker and a pure-tone signal, both centered in the region of 500 Hz, in this case 498 Hz. All stimuli were generated prior to the experiment, with 100 independent masker tokens and one token each of a brief signal and a long duration signal. The temporal and spectral characteristics of the stimuli were manipulated by way of convolution in the time domain with a bandpass filter response. The convolution array was 512 points, defined in the frequency domain as an array of zeros with a single point set to a value of 1 to define the passband. After transforming this array into the time domain, a Hanning window was applied. The sampling rate was 10 kHz. Figure 4 illustrates the effects of convolving an impulse with this array. The stimulus depicted, which served as the brief signal, is approximately 58 Hz wide at the 10-dB down points, and is effectively 45 ms in duration. Masker tokens were generated by convolving 409-ms Gaussian noise samples by this array, a process that introduced ramps and produced average long-term power spectra similar to that illustrated in Fig. 4. The long-duration signal was generated by convolution with a 498-Hz pure tone; this stimulus differed from the masker and the brief signal in that there was a prominent peak in the power spectrum at the signal carrier frequency, but onset and offset ramps were the same as those illustrated in Fig. 4.

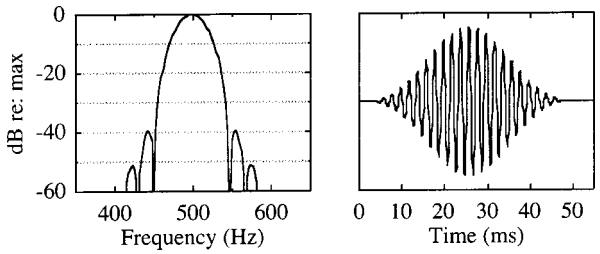


FIG. 4. The effect of convolution on an impulse is shown in the frequency domain (left) and in the time domain (right). This stimulus served as the brief signal for experiment 2.

Because the spectral density of these stimuli was not rectangular, it was decided to describe masker level in dB SPL, in contrast to the units of dB spectrum level used in experiment 1. Four masker levels were tested: 25, 40, 55, and 70 dB SPL.

3. Procedures

The procedures used in experiment 2 were the same as those in experiment 1 with the following exception. Masker tokens were randomly selected from among 100 tokens on each presentation interval. These were retrieved from disk and played out at 10 kHz, with an antialiasing filter at 3 kHz (Kemo, VBF 8). For those observers who participated in both experiments, experiment 2 was begun approximately 1.5 years after completion of experiment 1.

B. Results

The results of experiment 2 are shown in the right two columns of Fig. 5, with plotting conventions following those used in Fig. 2. To facilitate comparison of results with experiment 1, data from that experiment for Obs 1 and Obs 6 are shown again in the left column. Masker level is indicated along the abscissa in units of dB SPL. When making com-

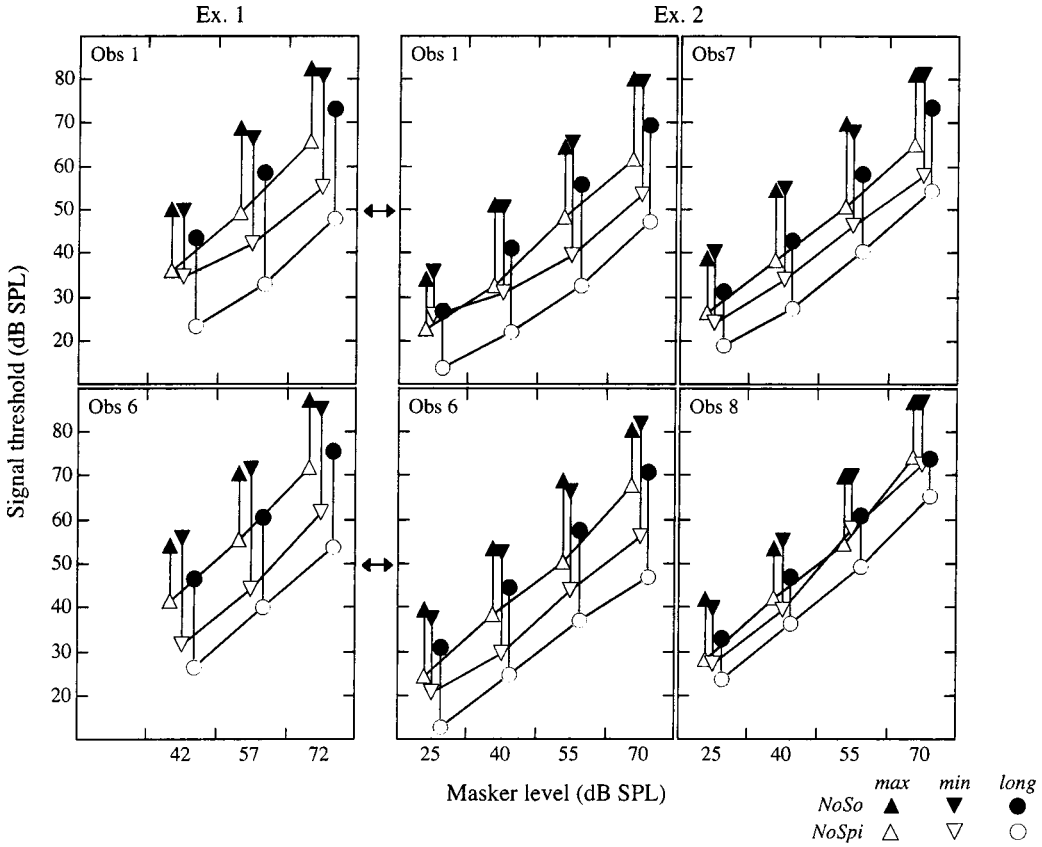


FIG. 5. Mean thresholds are plotted as a function of masker level in dB SPL. The right two columns show data from experiment 2, and the left column shows data from experiment 1 for Obs 1 and Obs 6, replotted for comparison.

parisons across data from the two experiments, it should be kept in mind that the levels used were not identical and that the temporal and spectral characteristics of the stimuli were quite different. Despite these differences the results for Obs 1 and Obs 6 are similar.

Thresholds from experiment 1 for the 42-, 57-, and 72-dB SPL masker levels (corresponding to 25, 40, and 55 dB spectrum level) are most closely comparable to those from experiment 2 at 40, 55, and 70 dB SPL, accepting the level mismatch of 2 dB. For Obs 1, in both experiments thresholds in the *min* condition were lower than those in the *max* condition at the highest level, but the difference was markedly reduced at the lower masker level. For Obs 6 in both experiments thresholds in the *min* condition were lower than those in the *max* condition consistently across these three masker levels. It is interesting to note that the *min* signal condition advantage was reduced at the 25 dB SPL masker level in experiment 2 for this observer.

The results of Obs 7 resemble those of Obs 6, though with a smaller overall MLD, while the results of Obs 8 are unique. For this observer thresholds in the *min* and *max* signal conditions are comparable at all masker levels. Another interesting aspect of this observer's data is the small MLD in the *long* condition. The *long* condition MLDs for Obs 8 were all below those obtained with any masker level for the other observers, and the lowest measured in experiment 2. This is consistent with the hypothesis that the MLD with a long-duration signal is driven by cues present in the epochs of higher signal-to-noise ratio; that is, this observer's small MLD for the long-duration signal could be due to failure to take advantage of cues present during the epochs of high signal-to-noise ratio associated with masker modulation minima, a possibility that is bolstered by findings of small values of the MLD in the *min* condition.

Overall, the findings of experiment 2 suggest that the pattern of results obtained in experiment 1 were not substantially affected by spectral cues. Further discussion of the results presented here will focus on those of experiment 1 because the signals used in that paradigm are briefer than those in experiment 2.

IV. DISCUSSION

Results of the present study show that the MLD is often considerably larger for brief signals coincident with epochs of low masker level (*min* condition) than epochs of high masker level (*max* condition). In some respects, this might appear inconsistent with the well-known finding that the MLD for a long-duration signal is *smaller* at low masker levels than at high masker levels (Blodgett *et al.*, 1962; McFadden, 1968). This apparent inconsistency can be resolved, however, by examining the effect of level on thresholds in the NoSo and NoSπ conditions. For both the *min* and *max* NoSo conditions, thresholds tended towards a constant signal-to-noise ratio across masker level. Thresholds in the *max* NoSπ condition likewise tended towards a constant signal-to-noise ratio. In contrast, in the *min* NoSπ condition, thresholds for most observers were at a lower signal-to-noise

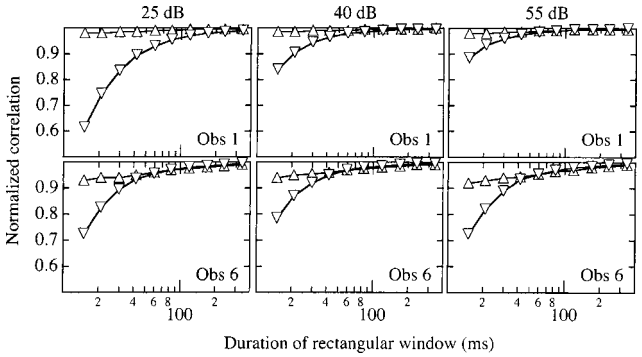


FIG. 6. Estimates of normalized correlation at threshold are plotted for the *min* (▽) and *max* (△) signal conditions based on data from Obs 1 and Obs 6 in experiment 1 for the 25-, 40-, and 55-dB spectrum level masker in panels 1–3.

ratio for the high masker level than for the low masker level. It has been argued that the dichotic cue available in the NoSπ condition allows the observer to ''listen in the dips,'' a process that is not possible in the baseline NoSo condition; as such, this cue produces a larger MLD for the *min* than for the *max* conditions. The level effect in the *min* NoSπ data results in a larger MLD at higher masker levels, analogous to the level effect noted for a long-duration signal and consistent with the hypothesis that thresholds for a long duration signal are ''driven by'' cues coincident with masker minima. The remainder of the discussion will focus on the level effects noted in the NoSπ condition.

Bernstein, Trahiotis, and colleagues (Bernstein *et al.*, 1999; Bernstein and Trahiotis, 1996) have suggested that a change in the interaural correlation of the stimuli can account for thresholds obtained in the NoSπ configuration. According to their model, performance in the dichotic conditions can be modeled by the following transformations: (1) exponential compression of the stimulus envelope, reflecting the nonlinear input/output characteristics of a normal-functioning cochlea, with exponent$=0.23$, to account for the active properties of the basilar membrane; (2) half-wave, square-law rectification, mimicking hair-cell transduction; (3) low-pass filtering, to account for the decreasing temporal synchrony of auditory-nerve fibers to stimulus fine structure as frequency is increased; and (4) calculation of the normalized correlation of the stimuli processed in this way. Because this model was developed for processing longer duration signals than those in the *min* and *max* signal conditions, this basic model was augmented here with a temporal window centered on the signal, inserted just prior to calculation of the normalized correlation. The output of the model is highly dependent on the duration of the temporal window, so a range of durations was assessed. A series of 500 stimuli generated as described in experiment 1 was run through the model incorporating a rectangular window centered on the signal, with one of ten durations ranging from 15 to 340 ms in equal log steps. The normalized correlation within the window for a signal at threshold forms the basis for the signal detection in this model, and results were compiled for each of the six observers' mean *min* and *max* NoSπ thresholds. Figure 6 shows an estimate of the normalized correlation plotted as a function of temporal window duration for two observers, Obs 1 and Obs 6 (top and bottom rows, respectively). The correlation for an No masker alone would be 1.0, so in this model cue strength is represented as the deviation from 1.0, with smaller values of correlation indicating greater cue strength. Left to right, columns show results for the 25-, 40-, and 55-dB spectrum level maskers. As in previous figures, results for the *min* condition are indicated with the down-pointing triangle and those for the *max* condition with the up-pointing triangle. Standard error of the mean around each estimate is smaller than the symbol in all cases.

The most salient feature of these results is the relatively low correlation for the very brief duration window in the *min* condition. This is due to the fact that a very brief window centered on a masker minimum limits the masker energy that is submitted to the calculation of correlation, such that addition of an Sπ signal has a large impact on the correlation. This advantage falls off with increases in window duration. While this general pattern of results is similar across masker levels, there is a discrepancy in the point at which signals in the *min* and *max* conditions are predicted to be equally detectable for some observers. Because signal level corresponded to threshold for each condition, correlation ought to be equal for the *min* and *max* conditions if it is to accurately predict performance. Based on visual inspection of Fig. 6, this would occur at different window durations for Obs 1 (ranging from 340 ms at 25 dB to 60 ms at 55 dB), while results for Obs 6 suggest greater consistency in the duration associated with equal cue strength (approx. 60 ms). This suggests that with this basic model no single rectangular temporal window accurately predicts average performance in the *min* and *max* conditions at all masker levels for Obs 1, but a single (relatively short) duration would accurately predict results for Obs 6. In order to quantify these observations, the point at which the *min* and *max* were associated with equal cue strength (i.e., equal correlation for a signal at threshold) was defined as the minimum window duration for which the difference between correlations was less than 0.03. The log transform of that duration for each observer at each of three masker levels was submitted to a repeated measures ANOVA.[1] This resulted in a significant effect of masker level ($F_{2,10}=11.33$, $p<0.005$) and a significant effect of observer ($F_{1,5}=325.93$, $p<0.0001$). This result supports the initial observations that the functions for the *min* and *max* conditions converge at different durations across masker level and that this result is highly variable across observers.

There are several hypotheses that might account for the observed level effects, including level-dependent compression, level effects on the binaural temporal window, and physiological noise. At the outset of the study it was hypothesized that compression might play a larger role in results obtained at the higher masker levels. Studies of basilar-membrane physiology have shown the input/output function to be more linear at low stimulus levels (e.g., Ruggero *et al.*, 1997). To the extent that compression plays a role in the detection advantage in the *min* conditions, this effect might be expected to decrease with decreasing stimulus level. Interestingly, the basic model predictions shown in Fig. 6 do not depend critically on the compression stage of the model.

The model was modified to incorporate a two-region compression function in place of the original exponential compression, with linear input/output up to 30 dB and a compression portion with a slope of 0.2 dB/dB thereafter, such as has been suggested based on physiological work (e.g, Ruggero *et al.*, 1997). The pattern of results with this modified model remained unchanged, and in fact omitting compression from the model altogether also failed to change the pattern of results, suggesting that level effects illustrated in Fig. 6 need not be attributed to compression. The failure to demonstrate a critical role for compression in this pattern of results could be due to the fact that the masker level is less variable during the signal epoch for these stimuli as compared to more typical MLD stimuli, so attenuation of masker maxima relative to minima has little effect over the interval of note (i.e., the brief epoch in the temporal center of the masker). This discussion, therefore, should not be taken as challenging the role of compression in the MLD in general, but rather as a suggestion that compression may not be a causative factor in the level effects observed here for brief signals coincident with masker envelope minima and maxima.

A second possible explanation for the level effects observed here is that the three masker levels are associated with different temporal window durations, and that the functions in Fig. 6 reflect constant cue detectability for different temporal windows. Characterizing the monaural temporal window, Moore *et al.* (1988) reported that lower stimulus levels are associated with longer temporal windows. In that study, a time constant of 10.8 ms was obtained based on data collected with a 30-dB stimulus, and a time constant of 8 ms was obtained with a 45-dB stimulus; that is, the temporal window duration increased by a factor of 1.35 for a 15-dB decrease in stimulus level. Though it is tempting to generalize this to binaural temporal windows, it should be pointed out that binaural temporal processing has been shown to differ importantly from monaural processing. Some studies of binaural hearing have produced significantly longer estimates of temporal resolution than those reported for monaural stimuli, though there is precedence for quite good binaural temporal resolution under some conditions (Bernstein *et al.*, 2001). Studies of temporal resolution have produced widely differing results, depending on the stimulus characteristics and demands task (Holube *et al.*, 1998), suggesting that multiple factors come into play in the binaural system (e.g., Kollmeir and Gilkey, 1990; Grantham, 1982). Finally, Culling and Summerfield (1998) reported that the binaural temporal window does not vary significantly with level, though there was a trend towards broader binaural temporal windows at lower stimulus levels in their data.

A third possible explanation for the masker level effect on the detection advantage associated with the *min* signal condition is the presence of physiological background noise (Dierks and Jeffress, 1962; McFadden, 1968). Yost (1988) reviewed the role of physiological noise in reduced MLDs obtained at low masker levels. As the masker level approaches that of the ongoing internal noise, thought to be partially correlated across ears, the interaural correlation of the masker-alone stimulus falls, and the decorrelation caused by introducing an Sπ signal becomes more difficult to detect (see van der Heijden and Trahiotis, 1997; Robinson and Jeffress, 1963). A similar effect could be responsible for the level effect observed in the present data set, in that the physiological noise would be present at a higher level relative to the masker for low masker levels, particularly during signal presentation epochs coincident with modulation minima.

Additional data were collected in a condition designed to simulate the hypothesized effect of physiological noise on thresholds in the 25-dB masker conditions by way of adding a background noise to the 55-dB masker. If physiological noise were responsible for the different pattern of results between the 55- and 25-dB masker, then thresholds in the 55-dB masker condition with the addition of a background noise stimulus should closely resemble the pattern of results obtained with the 25-dB masker. That is, the addition of background noise at an appropriate level should have little effect on thresholds in the NoSo conditions and in the *max* NoSπ condition, but should elevate thresholds in the NoSπ *min* condition. The background masker used here was generated digitally as the mix of a diotic source (No) and two uncorrelated sources (Nu) of Gaussian noise, with power of the Nu sources 2.33 times that of the No source; the resulting noise streams have a correlation of 0.3 (see Yost, 1988). Each stream was composed of a repeating 3.28-s token, and mixing was achieved digitally. The result was played continuously out of two channels of a DAC and added to the left and right stimulus channels, with stimuli and procedures otherwise identical to those of experiment 1, described above.

The data using the simulated physiological noise were only partially consistent with the idea that physiological noise is responsible for the level effect observed in the data of experiment 1. An initial level for the background noise masker was based on measurements of physiological noise with insert earphones collected by Berger and Kerivan (1983). While that study measured noise only up to 250 Hz, the pattern of those data suggests that a reasonable estimate of physiological noise levels at 500 Hz might be -10- to -5-dB spectrum level. Physiological noise at that level would be 30 to 35 dB below that of the 25-dB spectrum level masker. To simulate the effect of this noise with a 55-dB masker, a background noise of 20- to 25-dB spectrum level would be needed. Initial data collected in the presence of a 25-dB background noise showed little effect on thresholds, with larger MLDs in the *min* than the *max* condition for both Obs 1 and Obs 6. Raising the level of the added noise to 30-dB spectrum level, however, produced results more closely resembling the pattern of results obtained with the 25-dB masker. For Obs 1, introduction of simulated physiological noise reduced the 55-dB masker MLDs in the *min* condition from 25.6 to 13.6, a value that more closely resembled the MLD of 15.1 dB obtained at the 25-dB masker level. For the *max* condition MLDs dropped from 16.6 to 15.5, consistent with the MLD of 14.1 dB obtained at the 25-dB masker level. As expected, these effects were dominated by changes in the NoSπ thresholds, with NoSo thresholds remaining relatively unchanged with addition of background noise. For Obs 6, addition of the 30-dB background noise had much less impact on the pattern of the MLDs obtained. The MLD in the *min* condition was 20.2 dB with

the simulated physiological noise, as compared to 24.5 dB for the 25-dB masker in the original data set. The MLD in the *max* condition was 12.0 dB with simulated physiological noise, as compared to 12.7 dB with the 25-dB masker in the original data set. Interestingly, this observer also showed very little effect of masker level in the original data from experiment 1, perhaps indicating a lower level of background noise or a listening strategy that is less sensitive to background noise. While these preliminary results are broadly consistent with the predicted effects of physiological noise, the level of background noise necessary to simulate the effect was somewhat higher than expected.

V. CONCLUSION

Basic results with a brief duration signal confirmed that thresholds in the NoSπ condition are lower if that signal coincides with a masker modulation minimum than if it coincides with a maximum. This produces values of the MLD that are larger for the *min* than the *max* condition, particularly at high masker levels. A second data set collected with stimuli designed to eliminate potential spectral cues provides evidence that the results seen in experiment 1 are not due to a spectral cue. These results are consistent with the hypothesis that the MLD for long-duration tones in narrow-band noise, such as the *long* condition tested here, is driven by signal energy coincident with epochs of increased signal-to-noise ratios, equivalent to the cue available in the *min* signal condition.

Data collected at the highest masker level in experiment 1 are consistent with the normalized correlation model proposed by Bernstein, Trahiotis, and colleagues. It is interesting to note that inclusion of compression in this model is not critical to the prediction of a level effect in the finding of larger MLDs in the *min* than *max* signal conditions. The interaction between masker level and the advantage associated with the *min* condition is broadly consistent with inclusion of internal (physiological) noise, or alternatively with the hypothesis of a less sharply tuned binaural temporal window with decreasing stimulus level. Given the relatively high level of background physiological noise necessary to predict these effects, it is plausible that both of these factors could be contributing to the effects observed here.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD Grant No. 5 RO1 DC00418. A previous version of this work was presented at the 141st Meeting of the Acoustical Society of America [E. Buss, J. W. Hall, II, and J. H. Grose, ''MLDs for signals placed in masker envelope minima and maxima,'' J. Acoust. Soc. Am. **109**, 2375–2376 (2001)]. We would like to thank the Associate Editor, Leslie Bernstein, and two anonymous reviewers for their comments and suggestions on this work.

[1]Log-transformed data were used following the Plack and Moore (1990) treatment of monaural temporal window data. This transformation is not critical to the results reported here, however. A second ANOVA using duration in ms also produced a significant effect of level and observer ($p < 0.05$).

Berg, B. (**1989**). ''Analysis of weights in multiple observation tasks,'' J. Acoust. Soc. Am. **86**, 1743–1746.
Berger, E. H., and Kerivan, J. E. (**1983**). ''Influence of physiological noise and the occlusion effect on the measurement of real-ear attenuation at threshold,'' J. Acoust. Soc. Am. **74**, 81–94.
Bernstein, L. R., and Trahiotis, C. (**1996**). ''The normalized correlation: Accounting for binaural detection across center frequency,'' J. Acoust. Soc. Am. **100**, 3774–3784.
Bernstein, L. R., Trahiotis, C., Akeroyd, M. A., and Hartung, K. (**2001**). ''Sensitivity to brief changes of interaural time and interaural intensity,'' J. Acoust. Soc. Am. **109**, 1604–1615.
Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). ''The normalized interaural correlation: Accounting for NoSpi thresholds obtained with Gaussian and low-noise masking noise,'' J. Acoust. Soc. Am. **106**, 870–876.
Blodgett, H. C., Jeffress, L. A., and Whitworth, R. H. (**1962**). ''Effect of noise at one ear on the masked threshold for tone at the other ear,'' J. Acoust. Soc. Am. **34**, 979–981.
Buus, S. (**1985**). ''Release from masking caused by envelope fluctuations,'' J. Acoust. Soc. Am. **78**, 1958–1965.
Buus, S., Zhang, L., and Florentine, M. (**1996**). ''Stimulus-driven time-varying weights for comodulation masking release,'' J. Acoust. Soc. Am. **99**, 2288–2297.
Culling, J. F., and Summerfield, Q. (**1998**). ''Measurements for the binaural temporal window using a detection task,'' J. Acoust. Soc. Am. **103**, 3540–3553.
Dierks, J. K., and Jeffress, L. A. (**1962**). ''Interaural phase and the absolute threshold for tone,'' J. Acoust. Soc. Am. **34**, 981–986.
Eddins, D. A., and Barber, L. E. (**1998**). ''The influence of stimulus envelope and fine structure on the binaural masking level difference,'' J. Acoust. Soc. Am. **103**, 2578–2589.
Grantham, D. W. (**1982**). ''Detectability of time-varying interaural correlation in narrow-band noise stimuli,'' J. Acoust. Soc. Am. **72**, 1178–1184.
Grose, J. H., and Hall, III, J. W. (**1998**). ''Masker fluctuation and the masking-level difference,'' J. Acoust. Soc. Am. **103**, 2590–2594.
Hall, III, J. W., Grose, J. H., and Hartmann, W. M. (**1998**). ''The masking-level difference in low-noise noise,'' J. Acoust. Soc. Am. **103**, 2573–2577.
Hartmann, W. M., and Pumplin, J. (**1988**). ''Noise power fluctuations and the masking sine signals,'' J. Acoust. Soc. Am. **83**, 2277–2289.
Hirsh, I. (**1948**). ''Influence of interaural phase on interaural summation and inhibition,'' J. Acoust. Soc. Am. **20**, 536–544.
Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). ''Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments,'' J. Acoust. Soc. Am. **104**, 2412–2425.
Kollmeier, B., and Gilkey, R. H. (**1990**). ''Binaural forward and backward masking: Evidence for sluggishness in binaural detection,'' J. Acoust. Soc. Am. **87**, 1709–1719.
McFadden, D. (**1968**). ''Masking-level differences determined with and without interaural disparities in masker intensity,'' J. Acoust. Soc. Am. **44**, 212–223.
Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.
Plack, C. J., and Moore, B. C. J. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87**, 2178–2187.
Robinson, D. E., and Jeffress, L. A. (**1963**). ''Effects of varying the interaural noise correlation on the detectability of tone signals,'' J. Acoust. Soc. Am. **35**, 1947–1952.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
van der Heijden, M., and Trahiotis, C. (**1997**). ''A new way to account for binaural detection as a function of interaural noise correlation,'' J. Acoust. Soc. Am. **101**, 1019–1022.
van de Par, S., and Kohlrausch, A. (**1998**). ''Diotic and dichotic detection using multiplied-noise maskers,'' J. Acoust. Soc. Am. **103**, 2100–2110.
Yost, W. A. (**1988**). ''The masking-level difference and overall masker level: Restating the internal noise hypothesis,'' J. Acoust. Soc. Am. **83**, 1517–1521.

Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user[a)]

Christopher J. Long[b)] and Donald K. Eddington
Speech and Hearing Sciences Program, Division of Health Sciences and Technology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, Research Laboratory of Electronics, Massachusetts Institute of Technology, 50 Vassar Street, Cambridge, Massachusetts 02139, Cochlear Implant Research Laboratory, Department of Otology and Laryngology, Harvard Medical School, and the Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114

H. Steven Colburn
Hearing Research Center, Biomedical Engineering Department, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

William M. Rabinowitz
Bose Corporation, The Mountain, Framingham, Massachusetts 01701

(Received 21 August 2002; revised 1 July 2003; accepted 2 July 2003)

Experiments were conducted with a single, bilateral cochlear implant user to examine interaural level and time-delay cues that putatively underlie the design and efficacy of bilateral implant systems. The subject's two implants were of different types but custom equipment allowed presentation of controlled bilateral stimuli, particularly those with specified interaural time difference (ITD) and interaural level difference (ILD) cues. A lateralization task was used to measure the effect of these cues on the perceived location of the sensations elicited. For trains of fixed-amplitude, biphasic current pulses at 100 pps, the subject demonstrated sensitivity to an ITD of 300 μs, providing evidence of access to binaural information. The choice of bilateral electrode pair greatly influenced ITD sensitivity, suggesting that electrode pairings are likely to be an important consideration in the effort to provide binaural advantages. The selection of bilateral electrode pairs showing sensitivity to ITD was partially aided by comparisons of the pitch elicited by individual electrodes in each ear (when stimulated alone with fixed-amplitude current pulses at 813 pps): specifically, interaural electrodes with similar pitches were more likely (but not certain) to show ITD sensitivity. Significant changes in lateral position occurred with specific electrode pairs. With five bilateral electrode pairs of 14 tested, ITDs of 300 and 600 μs moved an auditory image significantly from right to left. With these same pairs, ILD changes of $\approx$11% of the dynamic range (in μApp) moved an auditory image from the far left to the far right—significantly farther than the nine pairs not showing significant ITD sensitivity. However, even these nine pairs did show response changes as a function of the interaural (or confounding monaural) level cue. Overall, insofar as the access to bilateral cues demonstrated herein generalizes to other subjects, it provides hope that the normal binaural advantages for speech recognition and sound localization can be made available to bilateral implant users. [DOI: 10.1121/1.1603765]

PACS numbers: 43.66.Pn, 43.66.Hg, 43.66.Ts [LRB]

I. INTRODUCTION

Approximately 60 000 people worldwide have a cochlear implant fitted to one ear. Although many of these monolateral cochlear implant users are able to achieve significant levels of speech understanding in quiet, they experience great difficulty understanding speech in noisy environments. People with a cochlear implant in each ear might benefit from the effects which are known to significantly assist normal-hearing listeners in understanding speech in noisy situations (Zurek, 1993) and improve sound localization. [For further information about binaural hearing phenomena, see reviews by Durlach and Colburn (1978), Grantham (1995), and Blauert (1997).]

Measurements of speech reception in conditions of spatially separated speech and noise sources show that some bilaterally implanted subjects score higher using both implants than in either monolateral condition, and that this benefit is predominantly based on attending to the better ear (van Hoesel and Clark, 1999; Gantz *et al.*, 2002; Muller *et al.*, 2002; Tyler *et al.*, 2002; van Hoesel *et al.*, 2002; van Hoesel and Tyler, 2003). In addition, bilateral cochlear implant users do show better localization of sound sources when using both implants (Gantz *et al.*, 2002; Tyler *et al.*, 2002; van Hoesel and Tyler, 2003).

[a)]Preliminary versions of this work were presented at the Conference on Implantable Auditory Prostheses, 19–24 August 2001, and the Conference on Binaural Hearing, Hearing Loss, Hearing Aids and Cochlear Implants, 22–24 June 2000, and in an unpublished thesis by Long (2000).

[b)]Author to whom correspondence should be addressed; MRC-CBU, 15 Chaucer Rd., Cambridge CB2 2EF, United Kingdom; electronic mail: stoph@alum.mit.edu

The main objective of our research was to characterize the binaural sensitivity of a subject with bilateral cochlear implants to basic bilateral stimuli. In particular, we examined interaural time difference (ITD) sensitivity using unmodulated, 100-pps trains of biphasic pulses. In our usage, *binaural sensitivity* refers to the ability of a subject to sense changes in two-ear stimuli that *cannot* be detected by using information from an analysis of separate left and right auditory images. Potential basic stimuli include those with interaural level differences (ILDs) and/or ITDs. Interpretation of sensitivity to ILD is, however, complicated by the fact that an ILD always involves a level change at one ear, raising the possibility that performance could be mediated by monaural cues. Even if overall level is roved to effectively eliminate the use of a single ear, the comparison of two monaural level estimates might allow localization judgments. In contrast, with measurement of ITD sensitivity no such confounding monaural sensitivity is present. We therefore used ITD as our primary assessment of binaural sensitivity and focused on the influence of ITD on the perceived lateral position of auditory images in binaural psychophysical experiments.

Previous studies of ITD sensitivity of bilateral cochlear implant users have revealed a wide range of performance, with the most sensitive subjects achieving thresholds of less than or equal to 150 μs (one subject in Pelizzone *et al.*, 1990; subject NU4 of Lawson *et al.*, 1996, 1998a; NU5 of Lawson *et al.*, 1998b; NU6, NU8, ME3, ME4, ME5, ME7, and ME8 of Lawson *et al.*, 2000; ME10 of Lawson *et al.*, 2001; A, B, C, and D of van Hoesel and Tyler, 2003), and some subjects requiring at least 1000 μs to detect an ITD (P1 of van Hoesel *et al.*, 1993; ME9 of Lawson *et al.*, 2001). Even within the same subject, performance can vary widely among ostensibly interaurally matched electrode pairs (e.g., from 50 to 450 μs with NU5 of Lawson *et al.*, 1998a).

Research with normal-hearing binaural listeners shows that matching the cochlear position of excitation across ears is necessary to obtain binaural sensitivity with acoustic stimuli. Henning (1974), Nuetzel (1976), and Nuetzel and Hafter (1981) explored cases that are relevant to bilateral cochlear implant research. Specifically, they examined the effects of interaural offsets in the carrier frequencies of sinusoidally amplitude-modulated, high-frequency tones on a subject's ability to discriminate ITDs in the modulator. The interaural offsets in carrier frequency are analogous to interaural offsets in electrode position. These studies showed that ITD sensitivity decreased as the narrow-band spectra at the two ears were separated. A similar study was undertaken by van Hoesel and Clark (1997) with two bilateral cochlear implant users, but these subjects showed only very poor sensitivity to ITD at best (1180 and 640 μs for P1 and P2, respectively), complicating interpretation.

We measured ITD sensitivity for 14 bilateral electrode pairs (spanning 8 mm) with a subject demonstrating good sensitivity to ITD ($\leqslant$600 μs) and evaluated binaural sensitivity as a function of the cochleotopic position of each pair's electrodes. The relation of ITD sensitivity to the results of direct pair-by-pair interaural comparisons of pitch (expanding upon the work of van Hoesel and Clark, 1997) was also examined, as were the ranges of ITD and ILD that influenced the perceived location of auditory images.

II. METHODS

A. Subject history

The 72-year-old female subject had binaural hearing as a child and first noticed her hearing loss at age 25. By age 44 she had little-to-no remaining hearing bilaterally. At age 59, her right ear was implanted with a six-electrode Ineraid cochlear implant, but she received only minimal benefit from the device (single-syllable word score $<$10%). At age 70 she received an eight-electrode pair Clarion cochlear implant (CI stimulator/receiver and Spiral electrode system) in her left ear with the hope that the newer technology would provide better performance. Unfortunately, this hope was not realized: her Clarion single-syllable word score was also below 10% (well below the 31% median performance of other users of this device in our laboratory). She briefly tried using both devices, but found this confusing and was using only the Clarion device on a daily basis when this study was conducted. Note that after losing her hearing, she spent at least 26 years as an adult without binaural hearing. Thus, she is in a class of listeners for whom a normal ability to connect binaural cues to source location was presumably established in childhood, but who lacked binaural input for many years.

B. Stimulus generation

Trains of fixed-amplitude, biphasic current pulses (76.9-μs/phase, 300-ms train duration) were delivered to either a single intracochlear electrode (monolateral) or to two intracochlear electrodes, one in each cochlea (bilateral/interaural). The repetition rate was 100 pps in the tests of lateral position and 813 pps in the pitch-ordering experiment.

A custom stimulation system (Long, 2000) allowed for close matching of the waveform parameters and fine control of ITD (jitter $<$0.5 μs) and ILD. Because the Clarion implanted current source is load dependent and the current cannot be directly measured after implantation, we report requested peak-to-peak current levels in quotes, i.e., "μApp."[1] The percutaneous connector of the Ineraid implant provides a direct, electric connection to each implanted electrode that allowed precise control of the current levels delivered to the electrodes. Thus, the stimulus levels for the Ineraid system are reported as μApp.

An ITD was introduced by delaying the stimulation of the Ineraid electrode relative to that of the Clarion electrode (Fig. 1).[2] Such a delay produces onset and ongoing interaural temporal differences.

Changes in ILD were produced by adjusting the Ineraid amplitude while holding the Clarion amplitude constant at a comfortable loudness level. Measurements are reported as a function of the Ineraid level, which can be interpreted in terms of ILD because the Clarion level was held constant.

All electrodes were stimulated in a monopolar configuration using an extracochlear return electrode. The medial electrodes of the Clarion array were used. For both electrode arrays, electrode 1 is the most apical. A schematic of the estimated electrode positions in the two cochleae is shown in

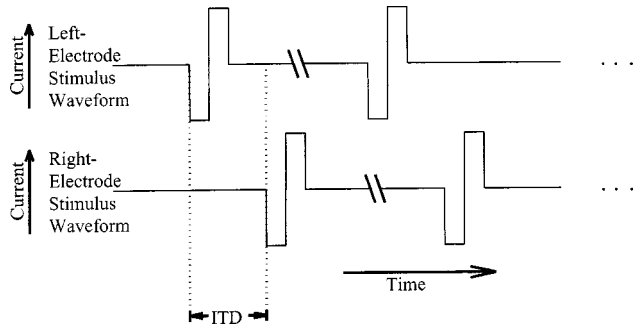


FIG. 1. Schematic illustration of the first two biphasic pulses in a bilateral train of pulses. Stimuli delivered to the right electrode (Ineraid) were delayed relative to those delivered to the left electrode (Clarion) to produce different ITDs. In addition, the amplitude of the right-electrode stimulus was varied in order to change the ILD.

Fig. 2. This schematic is based on the pitch-ordering results and is consistent with CT scans, both discussed later. In the case of bilateral stimulation, we refer to a ''bilateral electrode pair'' such as 7MC/2I (the 7th medial Clarion electrode and 2nd Ineraid electrode), which is actually a pair of electrode pairs, i.e., each pair is composed of an intracochlear electrode and a far-field ground.

C. Pitch-ordering test

A pitch-ordering test was used to rank order the electrodes by the pitch sensations they produced when stimulated individually and to identify bilateral electrode pairs that were closest in pitch. For each presentation, biphasic pulse trains (300 ms, 813 pps, loudness balanced) were delivered to each electrode of a bilateral pair in sequence (left ear first) and the subject identified which of the two stimuli was higher in pitch. The rate of 813 pps was chosen to minimize the impact of periodicity pitch (Eddington *et al.*, 1978b) and emphasize those characteristics of the elicited sensations that are determined by cochleotopic position. For example, Lawson *et al.* (1998a) reported better monaural electrode discrimination at 800 pps than at 200 pps (i.e., at 200 pps, they found 11 pitch-discriminable channels of stimulation on the left side and 13 on the right; and at 800 pps 13 on the left and 16 on the right). In addition to pitch measures we used CT scans to determine the relative placement of electrodes across sides (Ketten *et al.*, 1998).

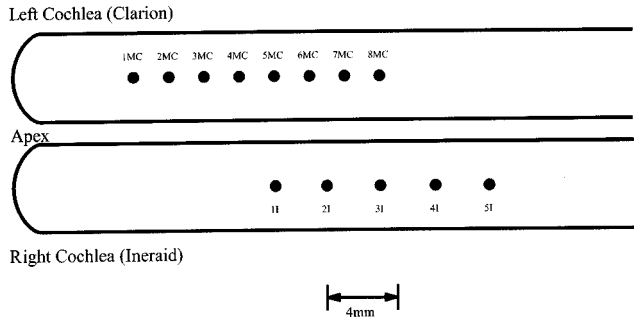


FIG. 2. Schematic illustration showing the relative longitudinal position of the two electrode arrays. Eight medial Clarion (1MC through 8MC) electrodes are shown in the left cochlea and five Ineraid (1I through 5I) electrodes are shown in the right. Locations shown are consistent with the CT scans and the pitch-ordering results. For both devices, electrode 1 is the most apical electrode. (The CT scans also showed that 6I was extracochlear.)

A run consisted of 25 presentations in succession to a single bilateral pair in the form of a two-interval (2I), two-alternative, forced-choice (2AFC) task. No feedback was provided. If the subject's first ten responses were identical, the run was terminated.

D. Preliminary tests

Each electrode's uncomfortable loudness stimulus level (I_{UCL}) was measured using a 11-2AFC (one-down, one-up, adaptive) procedure to determine the monolateral stimulus amplitude that resulted in a sensation level that was just uncomfortable. An electrode's absolute threshold level ($I_{\mathrm{threshold}}$) was measured using a 3I-3AFC (two-down, one-up, adaptive) procedure. $I_{\mathrm{threshold}}$ and I_{UCL} were determined for the standard trains of biphasic pulses at 813 pps. In addition, estimates (using a manual, bracketing procedure) were obtained at 100 pps for Ineraid electrodes 11, 21, and 31. These estimates were in good agreement with the other threshold/UCL measures and were used to calculate the dynamic range for each electrode and express stimulus levels relative to this range. Before measuring lateral position as a function of binaural cues, we determined reference levels for each bilateral electrode pair. Specifically, a centering test was conducted using bilaterally simultaneous, pulsatile stimulation (100-pps rate, 0-μs ITD). The Clarion stimulus amplitude was held at a comfortable loudness level and a 11-2AFC (one-up, one-down, adaptive) procedure was used to identify the Ineraid stimulus level that produced a fused sensation centered in the subject's head. After each presentation, the subject was to identify whether the auditory image was on the left or right. Such ''centered levels'' were identified for all but one of the bilateral electrode pairs tested. (Because the subject was unable to fuse the sensation elicited by stimulating 3MC/2I, a centered level was not identified and this pair was not included in the lateral-position test results.)

E. Lateral-position test

The perceived lateral position of a sensation elicited by bilateral stimulation was measured as a function of ILD and ITD for each of 14 bilateral electrode pairs. After each stimulus presentation, the subject assigned a number from a lateralization scale to represent the lateral position of a given sound sensation. The scale ranged from 1 to 7, where 1 represented a sound sensation at her left ear, 4 a sensation centered in her head, and 7 a sound sensation at her right ear. The subject responded ''none of the above'' (NA) if the sensation could not be characterized by the lateralization scale (e.g., if she heard more than one sound sensation). A stimulus set was made up of 15 bilateral stimuli with five ILDs selected to elicit sensations at locations covering the range from the left ear to the right ear (± 50 μApp; roughly ± 0.8 dB or $\pm 11\%$ of the dynamic range in μApp) and three ITDs (0, 300, and 600 μs) chosen to span the range of ITDs relevant in real-world situations. For each bilateral electrode pair, a total of 25 sets of 15 stimuli in random order were

TABLE I. Percent of presentations for which stimulation of the Clarion electrode produced a sensation that was judged higher in pitch than that elicited by the Ineraid electrode with both devices stimulated at 813 pps. The number of stimulus presentations for each electrode pair tested in shown is parentheses. The four scores classified as pitch indiscriminable (see text) are shown in **bold**. For both devices, lower electrode numbers correspond to more apical electrodes.

Clarion→ electrode	3MC	4MC	5MC	6MC	7MC	8MC
Ineraid ↓ electrode						
1I	0 (20)	19 (75)	**54 (100)**	72 (75)	68 (50)	76 (50)
2I			14 (35)	**48 (50)**	**42 (50)**	70 (50)
3I		0 (50)	24 (50)	16 (50)	16 (25)	**36 (75)**
4I						
5I		0 (10)				

presented. The bilateral electrode pairs were tested in the following order: 7MC/2I, 6MC/2I, 4MC/2I, 8MC/2I, 6MC/1I, 5MC/1I, 7MC/1I, 8MC/3I, 7MC/3I, 6MC/3I, 5MC/3I, 4MC/1I, 3MC/1I, 8MC/1I.

III. RESULTS

A. Pitch ordering

For each Ineraid electrode, the percentage of trials in which each Clarion electrode was judged higher in pitch was obtained and used to determine which bilateral pairs of electrodes could not be discriminated on the basis of pitch at 813 pps, as shown in Table I. Electrode pairs in which the percentages did not differ significantly ($p<0.01$) from 50% chance performance are shown in **bold** and are considered to produce approximately similar pitches. When scanning from left to right in a given row in Table I, the percentages generally increase. This is consistent with the cochleotopy of the electrodes because it indicates the pitch associated with a Clarion electrode tends to increase as one moves basally (i.e., toward 8MC). (Scanning from bottom to top in a given column shows the analogous effect for the Ineraid.) Bilateral electrode pairs 5MC/1I, 6MC/2I, 7MC/2I, and 8MC/3I were classified as pitch indiscriminable and were considered good candidate pairs for examining ITD sensitivity.

The relative positions of the two electrode arrays as deduced from the pitch matches were also compared to relative position estimates from CT scans, which were examined by Dr. Darlene Ketten. Dr. Ketten examined the CT scans without knowledge of the outcomes of the pitch comparisons. Her analysis matched the pitch results: 1I was closest to 5MC, 2I was between 6MC and 7MC, and 3I was closest to 8MC.

B. Lateral-position judgments

Lateralization judgments are shown for the pair with the largest ITD effect (8MC/2I) alone in Fig. 3 and for all tested electrode pairs in Fig. 4. The mean lateral position response is plotted as a function of the level of the Ineraid stimulus with ITD as a parameter equal to 0, 300, or 600 μs. The NA responses, which constituted less than 10% of all responses, were considered separately.

Before considering the pattern over all electrode pairs, the results for the specific bilateral electrode pair 8MC/2I (which showed the largest effect of ITD of any pair), shown in Fig. 3, are discussed first. When the ITD$=0$ μs (solid line) and the centered level stimulus (658 μApp as identified by the vertical line) is delivered to the right electrode simultaneously with the 800-"μApp" stimulus delivered to the left electrode, the mean lateral position is 4 (centered). Increasing the right-electrode stimulus amplitude causes the mean lateral position to move toward the subject's right. Lateral position also depends on ITD: At the centered level, the mean lateral position shifts toward the subject's left when the right-electrode stimulus is delayed. For a right-electrode stimulus amplitude of 683 μApp, the position is pulled rightward, an ITD of 300 μs then shifts the position toward the left and that of 600 μs shifts it farther. All these shifts are significant ($p<0.01$ on a t-test) and demonstrate ITD sensitivity at these amplitudes.

Note that a significant ITD effect is not observed at all levels. For right-electrode stimulus levels of 608 and 633 μApp, the ILD alone causes the percept to be fully lateralized to the left ear, leaving no room for an additional leftward shift with increasing ITD. For a right-electrode stimulus level of 708 μApp, the lack of a significant ITD effect is

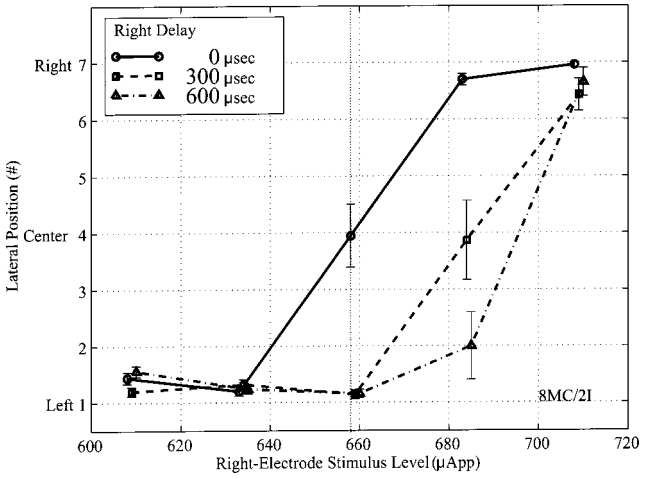


FIG. 3. Mean (±standard error) lateral position judgments plotted as a function of ILD and ITD for bilateral electrode pair 8MC/2I. Electrode 8MC was held at 800"μApp," while the amplitude of 2I was varied to create ILDs. The vertical line at 658 μApp indicates the 2I stimulus amplitude identified by the centering test (Sec. II D). Nonzero ITD data are shown slightly offset in amplitude for easier visual comparison, although these stimuli were presented at the same amplitudes as the zero ITD stimuli.

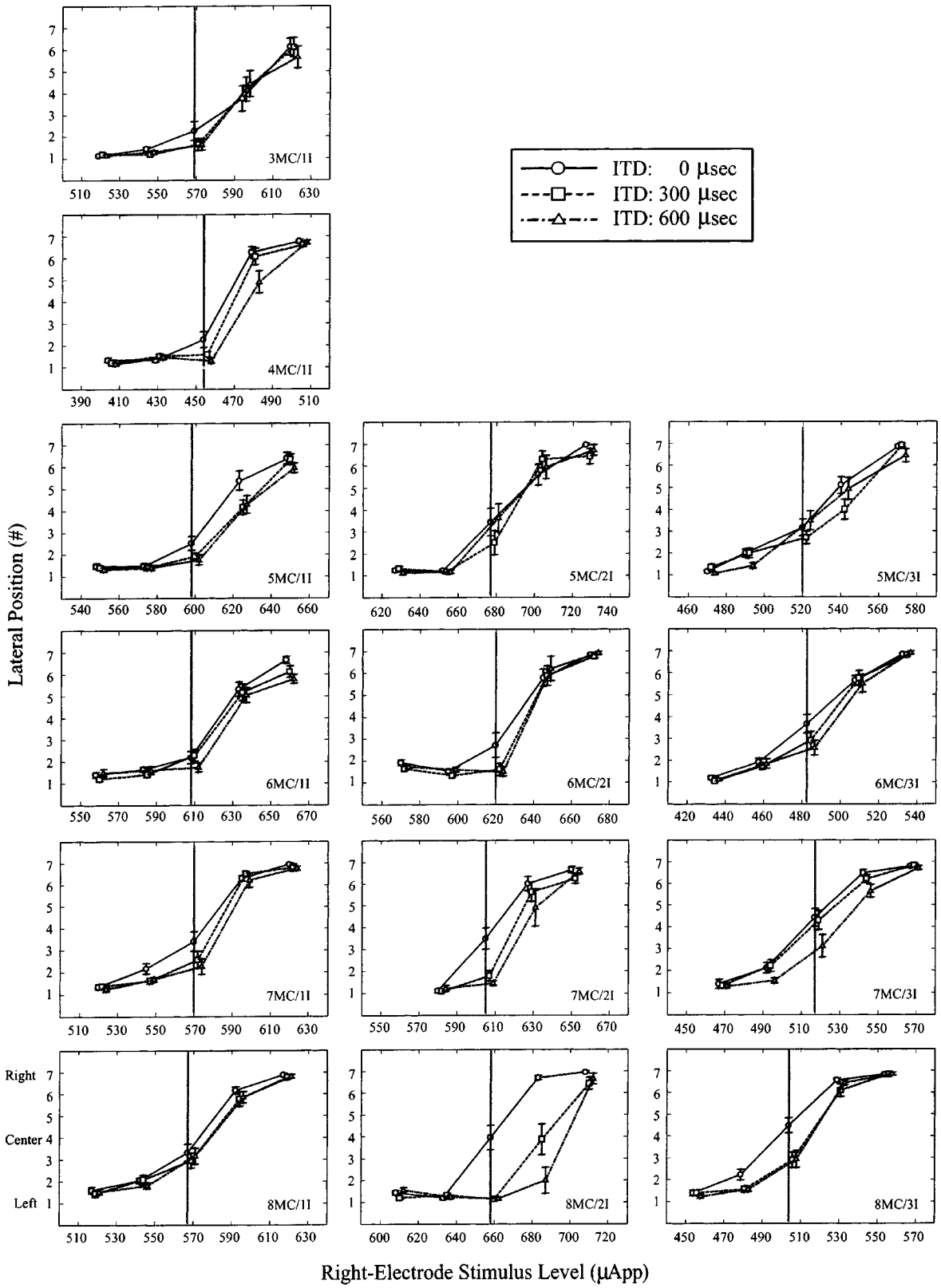


FIG. 4. Lateral position for 14 different electrode pairs plotted as a function of ILD with ITD as a parameter. Each panel represents the results for a single bilateral electrode pair. The panels are organized so that columns include data from a given Ineraid electrode and the rows for a given Clarion electrode. The electrode pair is identified in the bottom right corner of each panel. The Clarion stimulus level was held constant at a comfortable loudness level for each bilateral electrode pair. The Ineraid centered level, which produced a centered sensation in the preliminary test, is marked by a vertical line. On average, the stimulus level steps are 25 μApp or 5.4% of the dynamic range (in μApp); roughly equivalent to steps of 5.4 dB acoustically (Sec. IV C). The error bars represent $\pm$ standard errors. Nonzero ITD data are shown slightly offset in stimulus level for easier visual comparison, although these stimuli were presented at the same stimulus levels as the zero ITD stimuli.

consistent with results in normal hearing, where it has been observed that large ILDs can greatly reduce the effect of ITDs on lateralization (Sayers, 1964; Domnitz and Colburn, 1977).

Turning to the effects of the choice of electrode pair, consider the measurements for the full set of electrode pairs in Fig. 4. Each panel represents the results for a single bilateral electrode pair. The panels are organized so that columns include data from a given Ineraid electrode and the rows for a given Clarion electrode. The electrode pair is identified in the bottom-right corner of each panel. For all pairs shown, ILD (or equivalently, monaural level change) has a consistent effect on the lateral position judgments. When the right-electrode stimulus level is low, the subject assigns a low number on the scale. Increasing the right-electrode level increases the lateral position number assigned by the subject. On average, the stimulus level steps are 25 μApp, or 5.4% of the dynamic range (in μApp) and this can be considered roughly equivalent to steps of 5.4 dB acoustically (Sec. IV C).

Some bilateral electrode pairs show a substantial sensitivity to ITD (for example, the pairs 5MC/1I, 8MC/2I, and 8MC/3I). The magnitude of the ITD effect depends on the electrode pair stimulated. For example, beginning with 8MC/2I at the bottom of the middle column, the ITD sensitivity decreases (the area between the constant ITD lines is reduced) as one moves upwards toward more apical Clarion electrodes. A similar trend is seen in the right column. For the left column, the ITD effect is largest for 4MC/1I and 5MC/1I and near zero for 3MC/1I and 8MC/1I.

The subject answered ''none of the above'' (NA) in an average of 8.4% of presentations across all 14 bilateral electrode pairs shown. From our discussions with the subject, we infer that many NA responses were due to the presence of two separate auditory images. Such responses occurred most often when the ITD and ILD cues were contradictory.

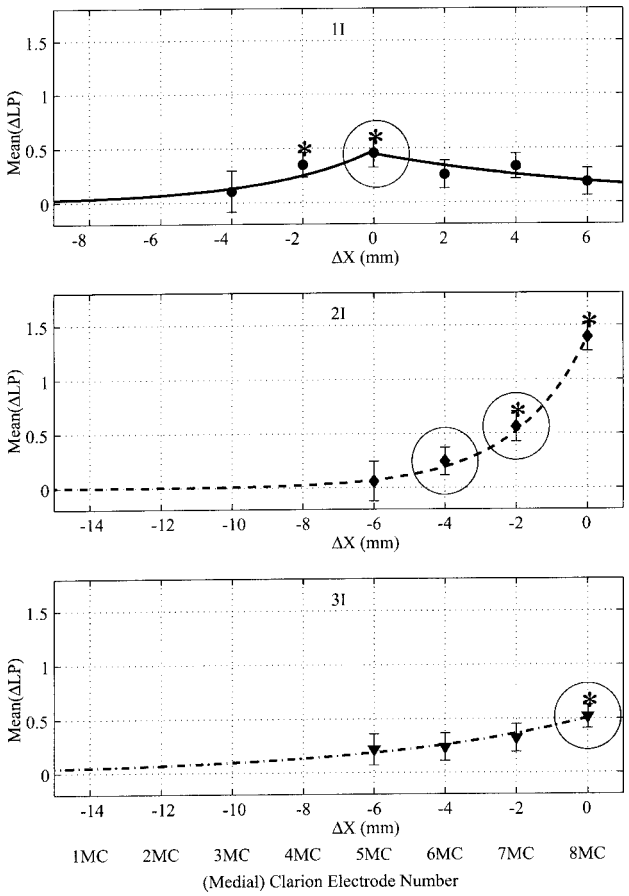


FIG. 5. Mean(ΔLP) as a function of the Clarion electrode, paired with three different Ineraid electrodes (as indicated at the top center of each panel). The Clarion electrode numbers are listed on the abscissa at the bottom of the figure. The ΔX values shown on the abscissa of each panel represent the difference in longitudinal position between a given Clarion electrode and the Clarion electrode exhibiting the greatest ITD sensitivity when paired with the Ineraid electrode of that panel. Asterisks identify mean(ΔLP) values significantly different from zero on a Bonferroni multiple comparison test (Rao, 1998) for $p<0.05$ and assuming 20 comparisons. Encircled values were found to be pitch indiscriminable (Table I). The curves are exponential fits to the data.

IV. DATA ANALYSIS AND DISCUSSION

A. Effect of ITD on lateral position

In order to directly compare the ITD sensitivity across all 14 bilateral electrode pairs, a metric was developed to characterize the effect of ITD on lateral position. The mean change in lateral position, mean(ΔLP), was computed for each bilateral electrode pair by adding the change in lateral position caused by ITD changes from 0 to 300 μs and from 0 to 600 μs at all ILDs and dividing by the number of differences. Specifically, for each bilateral electrode pair let

$\overline{\mathrm{LP}}_{j,k}$ = mean lateral position of the sensation measured for ITD j and right-electrode stimulus level k.

Then

$$\mathrm{mean}(\Delta\mathrm{LP}) = \frac{1}{2\mathrm{N}}\sum_{k=1}^{N}\left(\overline{\mathrm{LP}}_{0,k}-\overline{\mathrm{LP}}_{300,k}\right) + \left(\overline{\mathrm{LP}}_{0,k}-\overline{\mathrm{LP}}_{600,k}\right).$$

This measure is largest when a bilateral electrode pair shows consistent effects of ITD at multiple ILDs. The mean(ΔLP) is plotted as a function of the Clarion electrodes (Fig. 5), with statistically significant sensitivity shown by asterisks. The circled points will be discussed later. Each panel in this figure includes data from a bilateral electrode pair including a single Ineraid electrode (indicated at the top of the panel) and a set of Clarion electrodes (identified on the abscissa at the bottom of the figure). In addition, the difference in longitudinal position (ΔX) between each Clarion electrode and that Clarion electrode exhibiting the greatest ITD sensitivity when paired with the given panel's Ineraid electrode is shown on each panel's abscissa. Electrode combination 8MC/2I (middle panel, at $\Delta X=0$ mm) exhibits the maximum ITD sensitivity. Notice that pair 7MC/2I (at $\Delta X=-2$ mm) exhibits reduced ITD sensitivity, but remains significantly different from zero. Note also that a pairing with a more apical Clarion electrode (6MC/2I shown at $\Delta X=-4$ mm) does not show ITD sensitivity significantly different from zero. For a more apical Ineraid electrode (1I, top panel), significant ITD sensitivity is found only when paired with more apical Clarion electrodes (4MC and 5MC shown

at $\Delta X = -2$ and 0 mm, respectively). In addition to the mean(ΔLP), analyses employing different underlying assumptions (e.g., restricting analysis to centered levels or requiring that the change in lateral position exceed a threshold) support the trends reported in this section (Long, 2000).

B. Dependence of ITD sensitivity on relative electrode position

The sensitivity to ITD as indicated by the mean change in lateral position, mean(ΔLP), varies systematically with bilateral electrode pair (Fig. 5). Of the 14 bilateral electrode pairs tested, only five—shown by asterisks—exhibited significant ITD sensitivity.

To further examine how the difference in interaural electrode position affects ITD sensitivity, we estimated the length of the region in the left cochlea capable of supporting significant ITD sensitivity when paired with a specific Ineraid (right) electrode. A lower bound on the length of the region of sensitivity to ITD (LRS) was estimated as the *smallest region including all* bilateral electrode pairs exhibiting significant sensitivity to ITDs$\leqslant 600$ μs, and an upper bound was estimated as the *largest region including only* pairs that exhibit significant sensitivity to ITDs$\leqslant 600$ μs. The lower bound of the LRS for Ineraid 1I was estimated to be 2 mm (the nominal distance from 4MC to 5MC) and the upper bound to be 6 mm (from 3MC to 6MC). In the cases of Ineraid electrodes 2I and 3I, our data represented only one side of each maximum, so we estimated the LRS on the apical side of 8MC and multiplied by 2. The total LRS for 2I has a lower bound of 4 mm and an upper bound of 8 mm and, for 3I, the total LRS has a lower bound of 0 mm and an upper bound of 4 mm. Thus, the average lower bound on the LRS is 2 mm and the average upper bound is 6 mm. Monaural forward-masking studies with cochlear implant users have shown a spread of excitation similar to these bounds, e.g., Cohen (2001) showed 60% masking over 7% to 15% of the cochlear length (3 to 5 mm) for monopolar stimulation. For reference, a 5-mm difference in cochlear position corresponds to a frequency separation of an octave (Greenwood, 1990).

C. ITD sensitivity and other estimates of relative electrode positions

The pitch judgments in Table I (or equivalently the CT scan results) can be used to estimate the relative positions of the electrodes in the left and right cochleae, and the electrode locations in Fig. 1 were chosen on this basis. When ITD sensitivity analysis is combined with this figure, a pictorial summary representing the electrode pairs demonstrating significant ITD sensitivity is created, as shown in Fig. 6. When Ineraid electrode 1I was paired with Clarion electrodes, only pairs 4MC/1I and 5MC/1I showed significant ITD sensitivity. Notice that 5MC is closest to the longitudinal position of Ineraid electrode 1I. In contrast, the longitudinal position of electrode 2I is closest to Clarion electrodes 6MC and 7MC (and the pitch of electrode 2I was indiscriminable from that of electrodes 6MC and 7MC), but when electrode 2I was paired with electrode 6MC, a significant ITD sensitivity was not found. Also, even though Ineraid 2I elicited lower pitch sensations (Table I) than electrode 8MC, this pair exhibited the maximum ITD sensitivity of the 14 pairs tested. Note, when more than one Clarion electrode exhibited ITD sensitivity per Ineraid electrode, these Clarion electrodes were always nearest neighbors. This means that only electrodes in a particular region of the left cochlea produce ITD sensitivity when paired with a given right electrode.

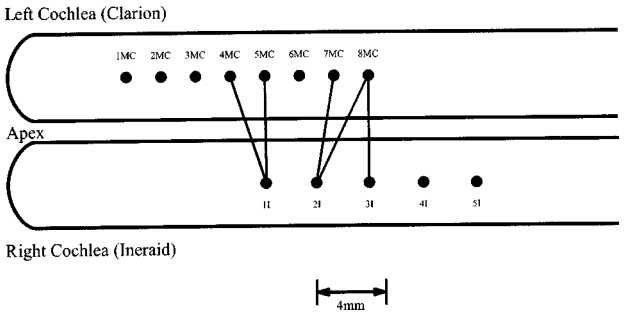


FIG. 6. Electrode pairs demonstrating a significant ITD sensitivity are linked by lines. Out of the 14 bilateral electrode pairs tested, significant ITD sensitivity was found for only the five indicated.

Of the four bilateral electrode pairs classified as pitch indiscriminable, two showed the greatest ITD sensitivity for their associated Ineraid electrode, one showed significant ITD sensitivity, and the last did not demonstrate significant ITD sensitivity. These trends are evident when one examines the mean(ΔLP) values of the pitch-indiscriminable bilateral electrode pairs shown encircled in Fig. 5. (The circles indicate the pitch-indiscriminable electrode pairs.) The mean $|\Delta X|$ for the four pitch-indiscriminable estimates is 1.5 mm [standard deviation (s.d.)=1.9 mm]. The mean $|\Delta X|$ of the other 20 bilateral electrode pairs available (three Ineraid electrodes with eight Clarion electrodes minus the four pitch-indiscriminable pairs) is 6.9 mm (s.d.=4.1 mm). These means are significantly different ($p<0.05$, t-test) and demonstrate that, on average, the pitch-indiscriminable electrode pairs are significantly closer to the most binaurally sensitive electrode position than the other bilateral electrode pairs. However, pitch-indiscriminable bilateral electrode pairs will not necessarily exhibit the maximum ITD sensitivity nor even significant ITD sensitivity. In fact, one pitch-indiscriminable bilateral electrode pair (6MC/2I) did not show significant ITD sensitivity and the best sensitivity was found with a pair that was pitch discriminable.

Our results show that interaural pitch comparisons do not allow exact predictions of the bilateral electrode pairs with significant or best ITD sensitivity. In addition to our analysis of our results, we calculated measures of binaural performance and pitch matching from the results of van Hoesel and Clark (1997). While the best ITD-JNDs the authors reported are quite large (1180 and 640 μs for P1 and P2, respectively) compared to the best ITD sensitivities reported, they do show a dependence on interaural electrode position difference. An exact agreement between the measures of pitch matching and binaural sensitivity (Fig. 2 of their paper) is found for only two of six bilateral electrode pairs. While the best ITD-JND was always within one step in cochlear distance (either 0.75 or 3.00 mm) of the pitch-matched elec-

trode pair, in one case for each subject (electrode pairs R13–L3 versus R13–L7 for subject P1 and R1–L10 versus R1–L14 for subject P2), the difference between the best ITD-JND and the ITD-JND with the pitch-matched pair is very large—a few hundred microseconds. Our data, together with these, indicate that the current practice of using a pitch-matching criterion is not sufficient to identify bilateral electrode pairs with maximum sensitivity to ITD. Selecting bilateral pairs where the place pitch of their individual members is indistinguishable does not guarantee significant ITD sensitivity (e.g., 6MC/2I), and bilateral pairs where the place pitch of their individual members is distinguishable may still exhibit significant ITD sensitivity (e.g., 8MC/2I was the pair with the largest mean(ΔLP)).

D. Accounting for differences in ITD sensitivity

The results showing better performance with electrode pairs that have comparable locations along the cochlear partitions are consistent with behavioral measures and with the anatomy of normally function auditory systems. Colburn's (1973) extension of the Jeffress model (1948, 1958) specified that the central binaural network only compares neural activity from interaurally corresponding (''tonotopically matched'') sets of auditory-nerve fibers from the left and right cochleae. This specification is consistent with physiological responses from cells in the medial superior olive (MSO), a reputed site for processing of ITD information, that respond preferentially when ipsilateral and contralateral stimuli have similar frequencies (Guinan *et al.*, 1972). Such a conceptual model leads one to consider several parameters of bilateral electric stimulation that may contribute to the degree of ITD sensitivity associated with a single, bilateral electrode pair. Selecting electrodes that excite segments of auditory-nerve fibers from the right and left that are matched and, therefore, elicit responses in higher-order neurons that are compared by the central network would seem to be an important factor in ITD sensitivity. To the extent the right and left excited fiber segments are matched, the number of neurons with correlated activity at the central network should be maximized. Support for the importance of the number of matched neurons is found in an analytical model of the MSO which predicts that the ITD-JND is inversely proportional to the square root of the number of activated MSO cells (Colburn and Isabelle, 1992).

If fibers associated with some cochleotopic segments have degenerated in one or both ears, the ITD sensitivity of bilateral electrode pairs that include an electrode positioned within such a segment may be severely impaired. It is also possible that increasing stimulus level may increase a bilateral pair's ITD sensitivity by not only increasing the number of active fibers within a given segment of the fiber array, but also by increasing the longitudinal extent of the nerve segment producing correlated responses.

The significant ITD sensitivities measured for bilateral pairs 4MC/1I, 5MC/1I, 7MC/2I, and 8MC/3I are consistent with the hypothesis that interaural electrodes of similar cochleotopic position are more likely to be sensitive to ITD. However, the highly sensitive pair (8MC/2I) and a pair not showing significant sensitivity (6MC/2I) contradict this notion and, therefore, warrant closer inspection.

The ITD sensitivity for 8MC/2I is much greater than that for the other bilateral electrode pairs tested. The conclusion that is most consistent with the data is that because 8MC and 2I were stimulated at high levels, the number of fibers eliciting responses that could be compared by the central network was larger than with other bilateral pairs. The amplitude of 8MC (800''μApp'') the patient labeled as comfortable when paired with 2I is the highest amplitude assigned to any Clarion electrode (group mean $=606$ ''μApp,'' s.d.$=86$ ''μApp''). Also, the current delivered to 2I (658 μApp) that produced a centered sensation is among the highest Ineraid amplitudes (group mean $=567$ μApp, s.d.$=61$ μApp). These relatively high stimulus levels may be one factor accounting for the high ITD sensitivity measured for an interaural electrode pair consisting of electrodes at rather different cochleotopic positions.

In contrast, bilateral electrode pair 6MC/2I does not show significant ITD sensitivity, even though the relative cochleotopic positions of 6MC and 2I shown in Fig. 6 make this result somewhat surprising. Because the pitch match between 6MC and 2I is better than that between 8MC and 2I (Table I), one might predict better ITD sensitivity with the former pair, but this is not the case. It is unlikely that stimulus level played a part in the poor ITD sensitivity when 6MC was paired with 2I as the Clarion comfortable loudness level for 6MC (610 ''μApp'') and the Ineraid centered level (620 μApp) are greater than or equal to the average amplitudes used (as noted in the discussion of 8MC/2I) and within the range used with pairings showing ITD effects. Thus, it appears the stimulus amplitudes used with this bilateral pair produced appropriate sensation levels. An alternative hypothesis is that fibers associated with the cochleotopic position of 6MC suffer some pathology that reduces the interaural correlation of their responses with responses elicited by 21. However, the threshold and UCL measures for 6MC, which one might expect to be elevated in the ''low-fiber density'' scenario, are not significantly different from those of other Clarion electrodes included in bilateral pairs exhibiting ITD sensitivity. In summary, it is not clear why pair 6MC/2I did not demonstrate significant ITD sensitivity. Whatever the cause, pitch comparisons were not sufficient to determine the absence of ITD sensitivity for this electrode pair. If 6MC/1I and 6MC/2I had shown significant ITD sensitivity the LRS lower and upper bounds would only increase by 2 mm.

E. Effect of ITD and ILD on lateral position

In this section we examine the relation between binaural cues and lateralization responses for all bilateral electrode pairs. The goals are to examine how ITD and ILD affect the perceived lateral position of sound sources and to determine the stimulus ranges that provide usable binaural information to bilateral cochlear implant users. The maximum change in lateral position associated with an ITD of 600 μs was calculated for each bilateral electrode pair. The changes in lateral position ranged from 0.25 to almost 4.75 lateral position units. The five bilateral electrode pairs that showed signifi-

cant ITD sensitivity (4MC/1I, 5MC/1I, 7MC/2I, 8MC/2I, and 8MC/3I) all show changes in lateral position greater than 1 lateral position unit.

In order to provide an intuitive measure of ILD and to allow relative comparisons across different electrode pairs, we express ILD as a percent of the dynamic range with a 100-pps stimulus. ''ILD(%DR)'' is defined as the difference between a right-electrode stimulus level (I_i) and the centered level (I_{centered}) expressed as a percent of the dynamic range of the right electrode

$$\text{ILD}(\%\text{DR}) = 100^*(I_i - I_{\text{centered}})/(I_{\text{UCL}} - I_{\text{threshold}}),$$

where the measurements of I_{centered}, I_{UCL}, and $I_{\text{threshold}}$ are described in Sec. II D.

The percent dynamic range measure was chosen with attention to the fact that the loudness growth of electric stimuli is roughly proportional to linear current, not to decibels of current, and that the entire dynamic range encompasses less than 10 dB on each electrode. Eddington *et al.* (1978a) measured the relation between acoustic and electric loudness with one of the few subjects having a cochlear implant and normal hearing at some frequencies contralaterally. Applying the equation they derived to our data, we find that increasing the stimulus level by 1% of the electric dynamic range is equivalent to increasing the SPL by 1 dB for an acoustic pure tone. Thus, we consider changes of 1% of the dynamic range (in μApp) roughly equivalent to steps of 1 dB acoustically.

In order to understand how stimulus levels influence binaural percepts, we calculated the change in lateral position (ΔLP) caused by a consistent ILD(%DR) for each bilateral electrode pair. Changing the ILD by 10.8% (s.d.$=1.1\%$) around the centered levels (from -5.5 to $+5.3\%$) caused changes in lateral position ranging from 2.3 to 5.5 lateral position units depending on the pair tested. In comparing the 14 pairs for which measurements were made, the five electrode pairs which showed ITD sensitivity were moved 4.7 (s.d.$=0.6$) lateral position units by this ILD, and the nine which did not show ITD sensitivity were moved 3.8 (s.d.$=0.7$) lateral position units—a significant difference ($p<0.05$, t-test). These results reflect a correspondence between the ITD sensitivity and the magnitude of the ILD effect. An ILD (or the associated, confounding monaural change in level) caused consistent changes in ΔLP with all electrode pairs tested. Future research should use a level rove to resolve if these changes are due to monaural or binaural processing.

V. SUMMARY AND CONCLUSIONS

We examined three general issues associated with the stimulation of single, bilateral electrode pairs in one cochlear implant user: (1) lateralization effects of binaural cues; (2) how binaural sensitivity depends on the bilateral electrode pair stimulated; and (3) the degree to which the pitch sensations elicited by individually stimulating members of a bilateral electrode pair predict the pair's sensitivity to ITD when stimulated simultaneously.

Our results demonstrate sensitivity to meaningful binaural cues in a single, bilateral cochlear implant user who suffered from a profound, long-term bilateral hearing impairment and who receives only a marginal benefit from either implant. The subject experienced fused sensations in response to stimulation of single interaural electrode pairs. The subject was sensitive to ITDs of 300 and 600 μs, ITDs within the range encountered normally. The sensitivity to ITD varied greatly across the interaural electrode pairs tested and showed a consistent pattern of monotonically decreasing sensitivity as the interaural pair's distance from the most sensitive pair increased. The magnitude of the ILD response was greatest with the five electrodes showing significant ITD sensitivity. While interaural pitch comparisons (and CT scans) showed some predictive value in selecting interaural electrode pairs that are sensitive to ITD, they were inadequate to reliably predict the pairs with maximum or even significant binaural sensitivity. Direct psychophysical measures of binaural sensitivity (e.g., ITD-JND) or physiological measures of binaural interactions (e.g., evoked potentials) will likely be required to accomplish this purpose.

Insofar as these results may be generalized, they provide hope that some measure of the normal binaural advantages for speech reception and sound localization can be made available to bilateral implant users if the interaural stimulus parameters like electrode pairing, interaural stimulus level, and interaural timing are appropriately optimized.

ACKNOWLEDGMENTS

We thank the subject who participated in the bilateral cochlear implant experiments. Her patience, hard work, and good humor were greatly appreciated. We are also grateful to Bob Carlyon, Richard van Hoesel, Les Bernstein, and an anonymous reviewer for their comments on a previous version of this manuscript, and to Darlene Ketten for analysis of the CT scans. This work was supported by NIH/NIDCD Training Grant T32 DC00038, NIH Contract N01-DC-6-2100, and a grant from the W.M. Keck Foundation.

[1]The manufacturer lists these requested current amplitudes as zero-to-peak values (Advanced Bionics Corporation, 1997).

[2]For notational convenience, a small fixed hardware delay ($2.19\ \mu$s $\pm 0.15\ \mu$s) is not included when reporting ITDs here.

Advanced Bionics Corporation (**1997**). CLARION Multi-Strategy Cochlear Implant: SCLIN Version 3.1 Supplement. Sylmar, California.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (Rev. ed.) (MIT Press, Cambridge, MA).

Cohen, L. T., Saunders, E., and Clark, G. M. (**2001**). ''Psychophysics of a prototype perimodiolar cochlear implant electrode array,'' Hear. Res. **155**(1–2), 63–81.

Colburn, H. S. (**1973**). ''Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination,'' J. Acoust. Soc. Am. **54**(6), 1458–1470.

Colburn, H. S., and Isabelle, S. K. (**1992**). ''Models of Binaural Processing Based on Neural Patterns in the Medical Superior Olive,'' in *Auditory Physiology and Perception: Proceedings of the 9th International Symposium on Hearing, on 9–14 June 1991*, edited by Y. Cazals, K. Horner, and L. Demany (Pergamon, Carcens, France), pp. 539–545.

Domnitz, R., and Colburn, H. (**1977**). ''Lateral position and interaural discrimination,'' J. Acoust. Soc. Am. **61**(6), 1586–1598.

Durlach, N., and Colburn, H. (**1978**). ''Binaural Phenomena,'' in *Handbook of Perception. Volume IV*: *Hearing*, edited by E. Carterette and M. Friedman (Academic, New York), pp. 365–466.

Eddington, D., Dobelle, W., Brackmann, D., Mladejovsky, M., and Parkin, J. (**1978a**). ''Auditory prostheses research with multiple channel intracochlear stimulation in man,'' Ann. Otol. Rhinol. Laryngol. **87**(6), 1–39.

Eddington, D., Dobelle, W., Brackmann, D., Mladejovsky, M., and Parkin, J. (**1978b**). ''Place and periodicity pitch by stimulation of multiple scala tympani electrodes in deaf volunteers,'' Trans. Am. Soc. Artif. Intern. Organs **24**, 1–5.

Gantz, B. J., Tyler, R. S., Rubinstein, J. T., Wolaver, A., Lowder, M., Abbas, P., Brown, C., Hughes, M., and Preece, J. P. (**2002**). ''Binaural cochlear implants placed during the same operation,'' Otol. Neurotol **23**(2), 169–180.

Grantham, D. (**1995**). ''Spatial Hearing and Related Phenomena,'' in *Hearing: Handbook of Perception and Cognition*, 2nd ed., edited by B. Moore (Academic, New York), pp. 297–345.

Greenwood, D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**(6), 2592–2605.

Guinan, Jr., J., Norris, B., and Guinan, S. (**1972**). ''Single auditory units in the superior olivary complex. II. Locations of unit categories and tonotopic organization,'' Int. J. Neurosci. **4**, 147–166.

Henning, G. (**1974**). ''Detectability of interaural delay in high-frequency complex waveforms,'' J. Acoust. Soc. Am. **55**(1), 84–90.

Jeffress, L. (**1948**). ''A place theory of sound localization,'' J. Comp. Physiol. Psychol. **41**, 35–39.

Jeffress, L. (**1958**). ''Medial geniculate body—a disavowal,'' J. Acoust. Soc. Am. **30**, 802–803.

Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., Gates, G. A., and Neely, J. G. (**1998**). ''*In vivo* measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays,'' Ann. Otol. Rhinol. Laryngol. Suppl. **175**, 1–16.

Lawson, D., Brill, S., Wolford, R., Wilson, B., and Schatzer, R. (**2000**). Ninth Quarterly Progress Report. 1 October through 31 December, 2000. NIH Project N01-DC-8-2105. Speech Processors for Auditory Prostheses. http://www.ninds.nih.gov/npp/.

Lawson, D., Wilson, B., Zerbi, M., and Finley, C. (**1996**). *Fifth Quarterly Progress Report. NIH contract N01-DC-5-2103*. http://www.ninds.nih.gov/npp/.

Lawson, D., Wilson, B., Zerbi, M., and Finley, C. (**1998a**). First Quarterly Progress Report. 30 September through 31 December, 1998. NIH Project N01-DC-8-2105. *Speech Processors for Auditory Prostheses*. http://www.ninds.nih.gov/npp/.

Lawson, D., Wilson, B., Zerbi, M., van den Honert, C., Finley, C., Farmer, Jr., J., McElveen, Jr., J., and Roush, P. (**1998b**). ''Bilateral cochlear implants controlled by a single speech processor,'' Am. J. Otol. **19**(6), 758–761.

Lawson, D., Wolford, R., Brill, S., Schatzer, R., and Wilson, B. (**2001**). Twelfth Quarterly Progress Report. July 1 through September 30, 2001. NIH Project N01-DC-8-2105. Speech Processors for Auditory Prostheses. http://www.ninds.nih.gov/npp/.

Long, C. (**2000**). ''Bilateral Cochlear Implants: Basic Psychophysics,'' Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Muller, J., Schon, F., and Helms, J. (**2002**). ''Speech understanding in quiet and noise in bilateral users of the MED-EL COMBI 40/40+ cochlear implant system,'' Ear Hear. **23**(3), 198–206.

Nuetzel, J. (**1976**). ''Spectral and Temporal Factors in the Lateralization of Complex Auditory Signals,'' Ph.D. dissertation, University of California, Berkeley.

Nuetzel, J., and Hafter, E. (**1981**). ''Discrimination of interaural delays in complex waveforms: Spectral effects,'' J. Acoust. Soc. Am. **69**(4), 1112–1118.

Pelizzone, M., Kasper, A., and Montandon, P. (**1990**). ''Binaural interaction in a cochlear implant patient,'' Hear. Res. **48**, 287–290.

Rao, P. (**1998**). *Statistical Research Methods in the Life Sciences* (Doxbory, Pacific Grove).

Sayers, B. (**1964**). ''Acoustic-image lateralization judgments with binaural tones,'' J. Acoust. Soc. Am. **36**, 923–926.

Tyler, R. S., Gantz, B. J., Rubinstein, J. T., Wilson, B. S., Parkinson, A. J., Wolaver, A., Preece, J. P., Witt, S., and Lowder, M. W. (**2002**). ''Three-month results with bilateral cochlear implants,'' Ear Hear. **23**(1 Suppl), 80S–89S.

van Hoesel, R. J. M., and Clark, G. (**1997**). ''Psychophysical studies with two binaural cochlear implant subjects,'' J. Acoust. Soc. Am. **102**(1), 495–507.

van Hoesel, R. J. M., and Clark, G. (**1999**). ''Speech results with a bilateral multichannel cochlear implant subject for spatially separated signal and noise,'' Aust. J. Audiol. **21**(1), 23–28.

van Hoesel, R. J. M., Ramsden, R., and O'Driscoll, M. (**2002**). ''Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user,'' Ear Hear. **23**(2), 137–149.

van Hoesel, R. J. M., Tong, Y., Hollow, R., and Clark, G. (**1993**). ''Psychophysical and speech perception studies: A case report on a binaural cochlear implant subject,'' J. Acoust. Soc. Am. **94**(6), 3178–3189.

van Hoesel, R. J. M., and Tyler, R. S. (**2003**). ''Speech perception, localization, and lateralization with bilateral cochlear implants,'' J. Acoust. Soc. Am. **113**(3), 1617–1630.

Zurek, P. M. (**1993**). ''Binaural Advantages and Directional Effects in Speech Intelligibility,'' in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by G. Studebaker and I. Hockberg (Allyn and Bacon, Boston), pp. 255–276.

The role of auditory feedback in sustaining vocal vibrato[a)]

Ciara Leydon, Jay J. Bauer, and Charles R. Larson[b)]
Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208

(Received 20 February 2003; revised 11 June 2003; accepted 30 June 2003)

Vocal vibrato and tremor are characterized by oscillations in voice fundamental frequency (F_0). These oscillations may be sustained by a control loop within the auditory system. One component of the control loop is the pitch-shift reflex (PSR). The PSR is a closed loop negative feedback reflex that is triggered in response to discrepancies between intended and perceived pitch with a latency of ~100 ms. Consecutive compensatory reflexive responses lead to oscillations in pitch every ~200 ms, resulting in ~5-Hz modulation of F_0. Pitch-shift reflexes were elicited experimentally in six subjects while they sustained /u/ vowels at a comfortable pitch and loudness. Auditory feedback was sinusoidally modulated at discrete integer frequencies (1 to 10 Hz) with ±25 cents amplitude. Modulated auditory feedback induced oscillations in voice F_0 output of all subjects at rates consistent with vocal vibrato and tremor. Transfer functions revealed peak gains at 4 to 7 Hz in all subjects, with an average peak gain at 5 Hz. These gains occurred in the modulation frequency region where the voice output and auditory feedback signals were in phase. A control loop in the auditory system may sustain vocal vibrato and tremorlike oscillations in voice F_0. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1603230]

PACS numbers: 43.70.Aj [AL]

I. INTRODUCTION

Vocal vibrato and vocal tremor are characterized by the superimposition of frequency and amplitude oscillations on the voice signal. Rhythmic oscillations have been observed in the activity of laryngeal (Hsaoi *et al.*, 1994; Koda and Ludlow, 1992; Niimi *et al.*, 1988), respiratory (Rothenberg *et al.*, 1988), and articulatory (Inbar and Eden, 1983; Sapir and Larson, 1993) muscles during vocal vibrato and tremor. The source of these oscillations is not well understood. It is believed that tremor may result from activity of a feedback control system associated with stretch receptors in muscles (Lippold, 1971), central oscillators, and ballistocardiograph, as well as from repetitive and asynchronous discharge of large motor units (Marsden, 1984). Winckel (1974) suggested that oscillatory muscle activity associated with vocal tremor and vocal vibrato arises from the activation of stretch receptors in laryngeal muscles. Recently, Titze *et al.* (2002) presented a model of vocal vibrato in which vibratolike oscillations in voice fundamental frequency (F_0) result from a negative feedback loop associated with stretch reflexes in pairs of antagonistic laryngeal muscles. While this model provides a persuasive explanation of how vocal vibrato is sustained, it does not account for a possible role for auditory feedback. Auditory feedback is known to play an important role in pitch and loudness control (Burnett *et al.*, 1998; Jones and Munhall, 2000; Kawahara and Williams, 1996; Lane and Tranel, 1971). In the present study, we propose that reflexes in the auditory system contribute to sustaining frequency and intensity modulations in the voice signal in vocal vibrato and vocal tremor.

Vocal vibrato is characterized by periodic fluctuations in F_0 and intensity in the singing voice. These pulsations typically occur at a rate of 4 to 7 Hz (Shipp and Izdebski, 1981) and with a F_0 modulation extent of ±1 semitone (100 cents). Vocal vibrato is a desirable feature of singing as it gives richness to a tone (Seashore, 1938), helps to separate a singer's voice from the orchestra (Sundberg, 1995), and makes vowels more prominent, allowing them to be more easily separated from background sounds (Marin and McAdams, 1991). The presence of vocal vibrato allows a listener to distinguish the voice of a trained singer from that of an untrained singer (Brown *et al.*, 2000).

Like vocal vibrato, vocal tremor is characterized by pulsations in F_0 and intensity (Koda and Ludlow, 1992). Pulsations in F_0 occur at similar rates in vocal tremor and vibrato. Ramig and Shipp (1987) observed that the average rate of F_0 oscillation was 6.8 Hz in vocal tremor and 5.5 Hz in vibrato. Shipp and Izdebski (1981) reported similar values with an average F_0 oscillation rate of 5 to 6 Hz in vocal tremor and of 5.4 Hz in vibrato.

In addition to the acoustic similarities outlined above, vocal vibrato and vocal tremor share a similar pattern of muscle activation. An electromyographic (EMG) study of intrinsic (thryroarytenoid, cricothryoid, and posterior cricoarytenoid) and extrinsic (sternothyroid and thyrohyoid) laryngeal muscle activity in patients with vocal tremor found spectral peaks at rates of 4 to 5.2 Hz (Koda and Ludlow, 1992). Sapir and Larson (1993) observed spectral peaks at similar frequencies in the EMG signals of anterior suprahyoid muscles (geniohyoid and genioglossus) and extralaryngeal muscles (thyrohyoid, strap, cricothyroid, and criocopharyngeal) in trained singers during vibrato. They re-

[a)]A portion of this material was presented in "Observations on the relation between auditory feedback and vocal vibrato" at the Third Biennial International Conference for Voice Physiology and Biomechanics, Denver, Colorado, September 2002.

[b)]Author to whom correspondence should be addressed. Electronic mail: clarson@northwestern.edu

ported that vibrato-related spectral peaks in EMG recordings occurred in these muscles at 4.83 Hz in a representative subject. Consistent with these observations, Hsaoi *et al.* (1994) reported modulations in cricothyroid activity as well as in thyroarytenoid muscle activity during vibrato.

The acoustic and EMG similarities between vocal vibrato and vocal tremor have led some researchers to speculate that these vocal features share common physiologic sources (Ramig and Shipp, 1987; Winckel, 1974). Ramig and Shipp (1987) speculated that a central generator produces tremor oscillations that are suppressed under normal circumstances. These oscillations are ''released from suppression by disease,'' resulting in a tremulous vocal quality. In contrast, these central oscillatory mechanisms can be ''selectively recruited'' to produce vocal vibrato (Ramig and Shipp, 1987, p. 166). Titze *et al.* (2002) proposed a model of vibrato in which oscillations in F_0 result from the presence of ''central oscillators'' and ''peripheral oscillators.'' The term ''central oscillators'' refers to neural input to muscles that originates at the cortical level and generates oscillations in planned muscle activation (Titze *et al.*, 2002). ''Peripheral oscillators'' refer to the biomechanical properties of the laryngeal system such as muscle mass and stiffness. These peripheral oscillators shape the oscillations in muscle activation generated by central oscillators (Titze, 1996; Titze *et al.*, 2002). Titze and colleagues (Titze *et al.*, 2002) described a ''reflex resonance model of vibrato'' in which a wide band of low-amplitude oscillations generated at the cortical level by central oscillations are shaped into a narrow band of oscillations through bi-directional negative feedback stretch reflexes. According to this model, the stretch reflex will generate oscillations at ~6 Hz.

Another factor that may contribute to the production of vocal vibrato is auditory feedback. Several studies have shown that auditory input influences the characteristics of vocal vibrato, suggesting that the auditory system may play a key role in regulating vibrato. Singers use auditory input to shape characteristics of vibrato such as rate and extent of modulation (Dejonckere, 1995; Deutsch and Clarkson, 1959; King and Horii, 1993; Vennard, 1967; Winckel, 1974). King and Horii (1993) reported that singers adjust their rate of vibrato to match a synthesized auditory signal presented at rates of 3 to 7 Hz. Deutsch and Clarkson (1959) observed that nonsingers decrease their rate and increase their extent of vibrato when producing vibrato under masking conditions. Also, Vennard (1967) noted that singers use auditory feedback to monitor the characteristics of their vibrato. When necessary, singers may fine-tune their vibrato by adjusting muscle activity. Winckel (1974) speculated that, during vibrato, the pitch and intensity of the voice signal might be adjusted continuously in response to auditory feedback. Thus, several studies support the suggestion that auditory feedback is capable of altering the characteristics of vocal vibrato.

Deutsch and Clarkson explored the relationship between the auditory system and vibrato oscillations (Deutsch and Clarkson, 1959; Clarkson and Deutsch, 1966). They proposed that vibrato results from a control loop within the auditory pathway that functions to maintain a steady note. Any unintended change in perceived pitch is corrected through compensatory adjustments in F_0. Corrections in F_0 are greater than necessary to return pitch to the desired level, causing the singer to systematically overshoot the desired pitch. Therefore, the singer must adjust pitch to compensate for the excessive correction. Again, the singer will overshoot the intended pitch. The singer will continue to adjust pitch towards the desired level and, in doing so, will overshoot the target pitch, resulting in continuous oscillations in pitch. According to Deutsch and Clarkson (1959), the frequency and extent of vibrato depends on the reaction time of the subject to an unintended change in pitch, the threshold necessary for an unintended pitch change to be detected and the rate at which an error can be corrected. Changing these factors will alter the characteristics of vibrato. They showed that delays in auditory feedback and decreases in the threshold for perception of a change in unintended pitch alter the frequency and amplitude of the frequency oscillations in untrained singers (Deutsch and Clarkson, 1959; Clarkson and Deutsch, 1966). However, in a review of the mechanisms underlying vocal vibrato, Sundberg (1987) noted that Shipp *et al.* (1984) questioned the role of auditory feedback in generating vocal vibrato in trained singers, as they reported that delayed auditory feedback (DAF) did not alter the characteristics of vocal vibrato. Given the conflicting findings regarding the effects of DAF on vocal vibrato, the role of auditory feedback in sustaining vibrato needs clarification.

In the present study, we sought to further demonstrate that properties of the auditory system sustain vibrato by experimentally eliciting pitch-shift reflex responses while subjects produced a steady tone. Previous studies revealed that subjects monitor their auditory feedback signal to regulate F_0 when producing steady vowels (Jones and Munhall, 2000), glissandos (Burnett and Larson, 2002), continuous speech (Donath *et al.*, 2002; Xu *et al.*, 2002), and when reading (Laukkanen, 1994). Any deviation in pitch from the intended level is corrected through a compensatory reflex, termed the pitch-shift reflex (PSR) (Burnett *et al.*, 1997). Like the stretch reflex, the PSR is a bi-directional negative feedback closed-loop reflex. When pitch is perceived to exceed the desired level, F_0 is decreased. Conversely, when pitch is perceived to be lower than the desired level, F_0 is increased. We propose that the PSR also contributes to the production of vocal vibrato.

Many studies investigating the influence of auditory feedback on voice production explored the vocal response to a single, transient shift in auditory feedback (Burnett *et al.*, 1997; Burnett and Larson, 2002; Donath *et al.*, 2002; Xu *et al.*, 2002). In the present study, we examined vocal responses to continuous sinusoidal modulation in auditory feedback. We hypothesized that responses to these modulations in auditory feedback (pitch-shift reflex) could sustain oscillations in voice F_0. We further speculated that the pitch-shift reflex (PSR) could sustain oscillations in F_0 with a period equal to twice the latency of the PSR. The latency of one PSR, causing either an increase or a decrease in F_0, ranges from 100 to 150 ms (Burnett *et al.*, 1998). This delay may represent the time needed for pitch decoding, error detection and motor command generation, transmission delays,

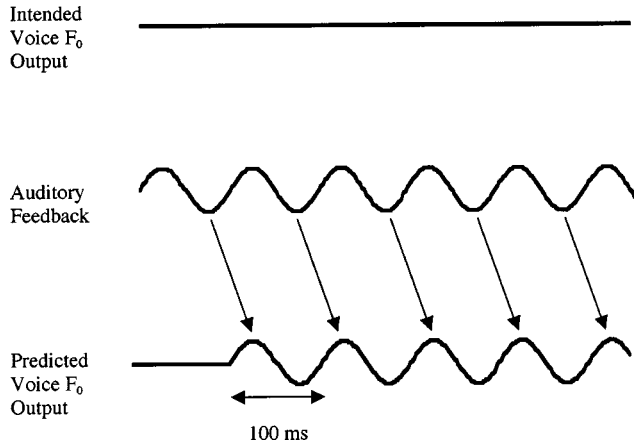


FIG. 1. Schematic illustration depicting the proposed manner in which the pitch-shift reflex can sustain vibratolike oscillations in voice fundamental frequency (F_0). The upper trace shows the intended F_0 contour (intended voice F_0 output). The middle trace shows perceived F_0 when real time auditory feedback is sinusoidally modulated at 5 Hz (auditory feedback). The lower trace depicts the expected F_0 signal in response to modulations in auditory feedback (predicted voice F_0 output). Decreases in perceived F_0 result in compensatory increases in voice F_0 with a 100-ms delay. Arrows indicate F_0 responses following perceived decreases in auditory feedback F_0. Conversely, increases in perceived F_0 result in compensatory decreases in voice F_0 also with a 100-ms delay. Arrows indicating voice F_0 responses corresponding to perceived increases in F_0 are omitted for clarity.

and muscle contraction speed (Kawahara and Williams, 1996). The delay for one complete F_0 cycle consisting of an increase and a decrease in pitch would correspond to two PSRs, or equivalently ~200 to 300 ms. Consecutive PSRs would generate cyclical oscillations in F_0 with a period of ~200 to 300 ms or, equivalently, a natural frequency of response of ~3.3 to 5 Hz (Fig. 1).

II. METHODS

A. Subjects

Six subjects (five females and one male; mean age 27.5 years) with normal hearing and no history of speech, language, voice, or neurological deficits participated in this study. None of the subjects were trained singers and all were naïve to the purpose of this study.

B. Stimuli

Subjects were seated in a sound-treated booth and instructed to repeatedly sustain the /u/ vowel sound at a comfortable and steady pitch for approximately 5-s intervals at 70 dB SPL at a 5-cm microphone-to-mouth distance (self-monitored visually with a Dorrough Loudness Monitor model 40-A). Vocalizations were limited to 5-s periods to ensure that subjects did not vocalize using their expiratory residual capacity. Vocalizations were transduced using an AKG boom-held microphone (model HSC 200). The signals were amplified with a Mackie Mixer (model 1202), and then processed for pitch shifting using an Eventide Ultraharmonizer (SE 3000). The pitch-shifted feedback signals were routed through HP decibel attenuators (model 350D), amplified to 80 dB SPL by a Crown audio amplifier (D75-A), and fed back to the subject in near-real time (8–20-ms delay) via over-the-ear AKG headphones (model HSC 200). Air-conducted auditory feedback was amplified by approximately 10 dB to mask potential bone conducted feedback. The above equipment was calibrated with a Brüel and Kjær 2203 sound level meter (weighting A).

Sinusoidal pitch-modulations were introduced in the auditory feedback of each subject at a single discrete integer frequency from 1 to 10 Hz. Auditory feedback was modulated at only one frequency per condition, yielding a total of ten experimental conditions per subject. The extent of each sinusoidal modulation of auditory feedback was ±25 cents (100 cents = 1 semitone), resulting in a peak-to-peak pitch modulation of 50 cents (0.5 semitones). The auditory feedback signal was modulated for the duration of each vocalization. A control condition was also collected for each subject, in which the auditory feedback signal was unmodulated. For each experimental condition, we analyzed voice output for a duration equivalent to 50 cycles of sinusoidally modulated auditory feedback. Thus, the rate of modulation of the auditory feedback signal determined the duration of the analysis window. For example, when a 1-Hz modulation was presented, 50 s of the voice signal were collected from 10 5-s vocalizations. Similarly, when a 10-Hz modulation of the auditory feedback signal was presented to the subject, 5 s of the voice signal were collected from one 5-s vocalization. As F_0 was not modulated under the control condition, 50 s of the voice output were collected across 10 5-s vocalizations.

C. Data analysis

Voice output and auditory feedback signals were digitally recorded at a sampling rate of 10 kHz (5-kHz anti-aliasing filter) on a laboratory computer using MacLab Chart v3.5 A/D conversion software (AD Instruments). In off-line analyses (Igor Pro software, version 4.0 by Wavemetrics, Inc.), these signals were low-pass filtered at 200 Hz for female and at 100 Hz for the male subjects, differentiated, and then smoothed with a five-point binomial, sliding window to remove high-frequency components from the audio signals. Voice F_0 was extracted from the preprocessed signals using a customized software algorithm in Igor Pro. The software algorithm detected positive-going threshold-voltage crossings, interpolated the time fraction between the two sample points that constituted each crossing, and calculated the reciprocal of the period defined by the center points. The resulting F_0 signals were converted to cents using the following equation: $\text{cents} = 100\,(39.86 \log_{10}(f2/f1))$, where f1 is an arbitrary reference note at 195.997 Hz (G4) and f2 is the voice signal in Hertz. The conversion of signals to cents permitted a comparison of the extent of change in the F_0 in response to a pitch modulation across all conditions regardless of baseline F_0.

A long-time average power spectrum of physiologic tremor was derived for each subject from the voice output signal using discrete Fourier transformations (DFTs) on vocalizations in which auditory feedback was not modulated. Long-time average spectra for the voice output and auditory feedback were also obtained for each modulation condition using DFT. A resonance curve was derived for each subject by measuring the amount of energy present at a given fre-

quency from 1 to 10 Hz following modulation at that frequency. Thus, the resonance curve is a composite curve obtained under ten experimental conditions.

Extent of oscillations in F_0 was determined for each subject at the modulation frequency (between 4 and 7 Hz) in which the greatest spectral peak was observed in the resonance curve. Voice F_0 output was band pass filtered between 1 and 20 Hz to remove extraneous noise from the waveforms, and averaged across a time window equivalent to two cycles of modulated feedback. Thus, duration of the averaging window varied according to the modulation frequency. The extent of F_0 oscillations was measured from the averaged waveform as the difference in cents between the highest amplitude and the lowest amplitude peaks. Extent of oscillation was expressed as mean, range, and standard error of the mean (SEM).

Transfer functions of the audio-vocal system were constructed for each subject by comparing the extent and phase of the voice output to the modulated auditory feedback signals at all frequencies (1 to 10 Hz). To obtain a transfer function of gain (dB), the extent of voice F_0 oscillations at a given frequency was divided by the extent of F_0 modulations in the auditory feedback signal at that frequency. Similarly, the phase relationship between the voice output signal and the modulated auditory feedback signal was determined by subtracting the phase of voice output F_0 oscillations from the phase of auditory feedback oscillations. A positive value would indicate that the voice output signal was phase-advanced relative to the modulated auditory feedback signal, while a negative value would suggest that the voice output signal was phase-lagged relative to the modulated auditory feedback signal. A zero value would indicate that the voice output and modulated auditory feedback signals were in-phase.

III. RESULTS

We compared the energy present in each subject's voice signal at frequencies from 1 to 10 Hz under experimental and control conditions. Figure 2 depicts fluctuations in voice F_0 obtained under experimental conditions in which auditory feedback was modulated at frequencies from 1 to 10 Hz (resonance curve) and under a control condition in which auditory feedback was not modulated (physiologic tremor) for a representative subject. While large amplitude peaks in the physiologic tremor spectrum obtained under the control conditions are absent, there are slight increases in amplitude at 2 Hz and at 6 to 8 Hz. The higher frequency peak (6 to 8 Hz) is within the range of F_0 modulation typically associated with normal tremor (Marsden, 1984). The resonance curve is a composite curve that depicts the amplitude of energy present at each frequency in response to sinusoidal modulation of feedback at those frequencies. The average resonance curve shows greater amplitude of energy, at frequencies ranging from 1 to 7 Hz, than is shown in the curve depicting physiologic tremor. A single peak in the average resonance curve occurred at 5 Hz, indicating that the greatest amplitude of energy was present at this frequency. Amplitude extent of modulation ranged from 8 to 16 cents across the six subjects (mean = 10.4, SEM = 3.5).

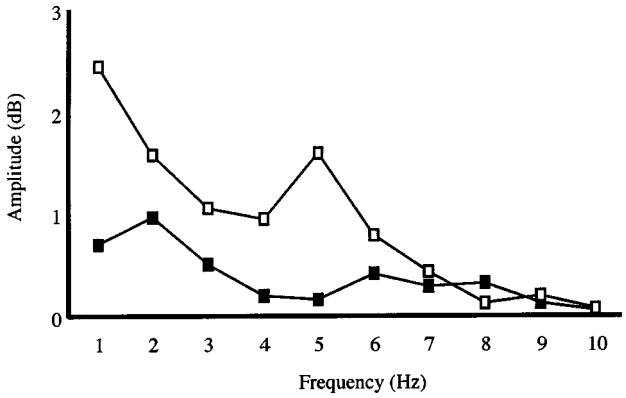


FIG. 2. Long-time power spectra for a representative subject. The physiologic tremor curve (solid squares) depicts tremor present in the voice signal in the absence of modulated auditory input. Observed spectral peaks at 2 Hz and at 6 to 8 Hz indicate increases in physiologic tremor. The resonance curve (open squares) reflects the composite energy present in the voice signal during modulated auditory feedback. This curve shows a spectral peak at 5 Hz indicating natural resonance of pitch-shift reflex at this frequency. The spectral peak at 1 to 2 Hz may represent a voluntary response to modulated auditory feedback.

Transfer functions for gain and phase are plotted for all subjects in Fig. 3. The transfer functions of gain (upper trace) show that greatest gains, or peaks, occurred at modulation frequencies within the vibrato frequency range (4 to 7 Hz) in all subjects. Although there were occasional peaks in the gain plots at low frequencies, these most likely reflected voluntary modulations in F_0. The transfer functions of phase (lower trace) show that the voice output and modulated auditory feedback signals were in-phase at 3 to 5 Hz across all subjects. The voice output signals were phase-advanced at low frequencies (2 Hz) and phase-lagged at mid to high frequencies (5 to 10 Hz) relative to the modulated auditory feedback signals for all subjects. Transfer functions of gain and phase averaged across all subjects (Fig. 4) illustrate the essential features of the gain and phase characteristics described above.

Time-aligned auditory feedback and voice output signals revealed a similar pattern. Figure 5 shows two cycles of time-aligned modulated auditory feedback (upper trace) and the corresponding voice output signal (lower trace) for one subject. The auditory feedback signal was phase-advanced at low frequencies (1–3 Hz), in-phase at mid frequencies (4 Hz), and phase-lagged at high frequencies (5–10 Hz).

IV. DISCUSSION

Our findings support the hypothesis that reflexes in the audio-vocal system can sustain vibratolike oscillations in voice F_0 in response to an imposed continuous sinusoidal modulation of the auditory feedback signal. Sinusoidal modulations of auditory feedback pitch elicited a pitch-shift reflex (PSR) resulting in oscillations in voice F_0 at frequencies close to twice the predicted latency of the PSR loop. Transfer functions for all subjects showed greatest gain in spectral energy of voice output between 4 and 7 Hz, with a mean peak gain at 5 Hz, indicating that the PSR elicited the greatest response to modulated auditory feedback at rates consistent with those typically observed in vocal vibrato.

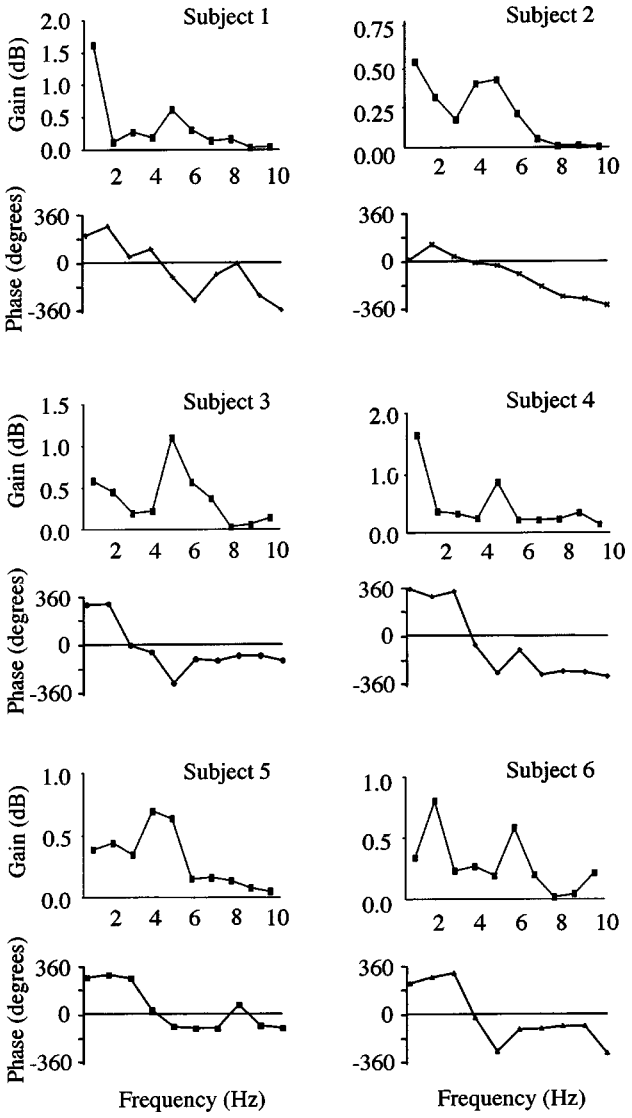


FIG. 3. Transfer functions of gain (upper trace) and of phase (lower trace) for all subjects. The transfer functions of gain depict the amount of energy present in the voice signal relative to the energy present in the modulated auditory signal. Peak gains occurred at frequencies ranging from 4 to 7 Hz indicating that the pitch-shift reflex demonstrated greatest gain at frequencies consistent with vibrato. Low-frequency peaks observed in the transfer functions of each subject may have resulted from voluntary responses to modulated auditory feedback. The transfer functions of phase show the phase relationship between voice output and modulated auditory feedback signals. Voice output and auditory feedback signals were in-phase at frequencies between 3 and 5 Hz.

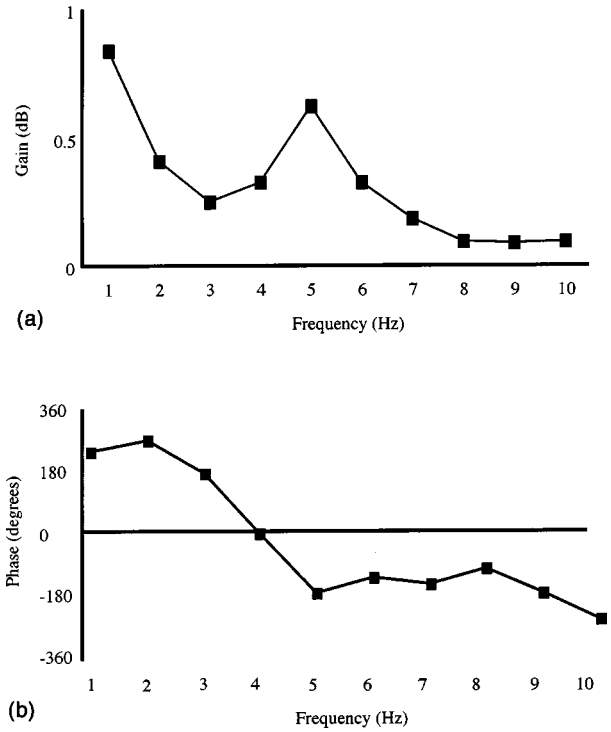


FIG. 4. (a) Average transfer function of gain. Peak gain occurred at 5 Hz, which is consistent with peak gain in vibrato. The low-frequency peak at 1 to 2 Hz may have resulted from voluntary adjustment to pitch in response to modulated auditory feedback. (b) Average transfer function of phase. Voice output and modulated auditory feedback signals were in-phase at 4 Hz.

Furthermore, the greatest gains were obtained at modulation frequencies where the modulated auditory feedback and the voice output signals were most in-phase. Our findings suggest that the PSR has a natural frequency that falls within the frequency range associated with vibrato and is consistent with the observed natural frequency of auditory response of 6.6 Hz observed by Kawahara and Williams (1996).

We observed low frequency peaks in transfer functions for all subjects (Fig. 3). The low frequency peaks may represent a voluntary adjustment to pitch in response to modulations in the auditory signal (Sapir and McClean, 1983). This interpretation is consistent with the hypothesis posited by Kawahara and Williams (1996) that pitch can be altered voluntarily at low frequencies in response to modulated auditory feedback. Alternatively, the low frequency peaks may reflect the effects of heartbeat on F_0 (Orlikoff and Baken, 1989). For these reasons, we do not consider these low frequency peaks in gain to result from the effects of PSR on voice F_0 and therefore we omitted them from further interpretation.

Sinusoidal modulation of auditory feedback elicited vibratolike oscillations in F_0 in the vocal output signal. However, the extent of F_0 oscillations in the voice output signal, as measured by the amplitude of energy at the spectral peak of modulated voice output, was lower than the 1–1.5 semitone range typical of vibrato. Extent of F_0 oscillations in the voice output signal following ± 25 cents feedback modulation did not exceed ± 8 cents in any subject. The relatively low extent of F_0 modulation observed in the present study suggests that the PSR determines the rate, not the extent of F_0 oscillations. This further suggests that rate, not extent, of F_0 oscillations may be controlled by reflexes in the auditory pathway. Based on our results, however, we cannot rule out the possibility that the extent of modulation of F_0 may have been depressed due to two factors. First, the extent of modulation of F_0 in vibrato is associated with the musical (Prame, 1995) and emotional (Seidner *et al.*, 1995) context of the piece. In the present study, the subjects were instructed to produce a steady pitch in a laboratory setting. Second, singers demonstrate more finely tuned oscillations associated with a greater spectral peak, and consequently greater extent, than nonsingers (Titze *et al.*, 2002). Had trained singers been

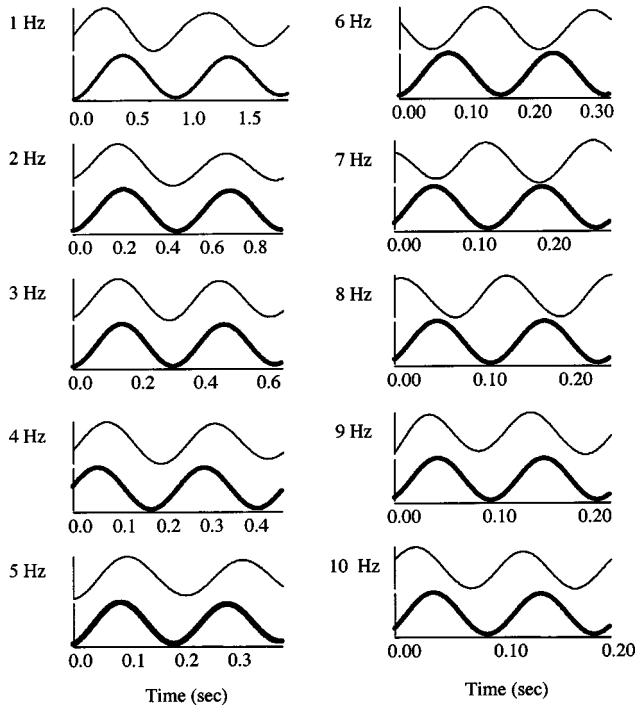


FIG. 5. Time-aligned voice output signal (upper trace) and modulated auditory feedback signal (lower trace) at each modulation frequency (1 to 10 Hz) for a representative subject. Two cycles of modulated auditory feedback are shown at each modulation frequency. The signals are most in-phase at modulation frequencies from 3 to 5 Hz. Greatest amplitude of response occurred when the voice output and auditory feedback signals were most in-phase. The large amplitude responses observed at 1 to 2 Hz may have been due to voluntary adjustments in response to modulations in auditory feedback.

asked to produce vocal vibrato, the extent of F_0 modulations may have been greater.

There was some variability in peak modulation frequency within the vocal vibrato range across subjects. On average, we observed greatest gain in voice output following modulation in auditory feedback at 5 Hz, with individual subjects' gains ranging from 4 to 7 Hz. Thus, while all subjects demonstrated greater energy at frequencies within the acceptable range for vocal vibrato, intersubject variability in frequency of peak gains was observed. This intersubject variability may have been attributable to differences in latency of the PSR across subjects. If a subject's PSR has a latency of ~100 ms, the peak in the voice modulated signal will be expected to occur at 5 Hz (the reciprocal of 200 ms). If the PSR is slower or faster than 100 ms, spectral peaks will be observed at a lower or higher frequency, respectively. Subjects may exhibit differences in latencies of the PSR due to nonuniform attendance to the pitch modulations (Burnett *et al.*, 1998; Larson *et al.*, 1996). Alternatively, subjects may use idiosyncratic strategies to produce the same vocal output (Sapir and Larson, 1993; Ramig and Shipp, 1987). That is, different muscle combinations and different levels of muscle activity may be used to produce vocal vibrato. As the viscoelastic properties of muscles vary, nonuniform responses to input from the auditory pathway are expected. Additionally, subjects may be able to modify the gain in their PSR loop by altering the characteristics of underlying neuromuscular mechanisms through task-dependent latency modulation (Hain *et al.*, 2000). Titze *et al.* (1994) hypothesized that the output of the central oscillators may be altered through training. At the peripheral level, singers may make adjustments to biomechanical properties of muscles (Sapir and Larson, 1993) which would alter the latency and magnitude of response of the muscles to input from the auditory reflex pathway and thereby increase or decrease the rate and magnitude of F_0 modulation. Finally, subjects may have intrinsic differences in the speed of neural transmission through the pathways associated with the PSR.

The extent of intensity modulation associated with variations in F_0 of the vocal output signal was not analyzed in the present study. It is possible that F_0 modulations imposed on the voice signal resulted in amplitude modulation in the auditory feedback signal due to resonance-harmonics interactions (Horii, 1989; Horii and Hata, 1988). That is, as a result of changes in F_0, the amplitude of the auditory feedback signal could have risen and fallen as harmonics moved closer and then further from the resonant frequencies of the vocal tract. Therefore, at present, we cannot rule out the possibility that subjects responded to changes in the intensity in addition to changes in the pitch of the auditory feedback signal.

Auditory feedback provides important information for fine-tuning voice output for singing. Deutsch and Clarkson (1959) suggested that a control loop within the auditory system maintains vibratolike oscillations in F_0. The findings of the present study are consistent with this hypothesis. We demonstrated that sinusoidal modulations in auditory feedback sustained oscillations in voice F_0 through consecutive compensatory reflexive responses. These oscillations occurred at approximately 5 Hz, which is the reciprocal of twice the latency of the pitch shift reflex. We propose that a negative feedback control loop within the auditory system, involving the PSR, drives F_0 oscillations associated with vocal vibrato. Given that vocal tremor shares many acoustic and EMG similarities with vibrato, we speculate that reflexes in the auditory pathway may play a role in sustaining vocal tremor during voicing in some individuals. Hain *et al.* (2001) hypothesized that abnormal timing in the auditory pathway may result in vocal tremor. Future studies should be directed at evaluating the role of the control loops within the auditory system in triggering or sustaining vocal tremor.

ACKNOWLEDGMENTS

This research was supported by NIH Grant No. DC02764-01 awarded to Dr. Charles R. Larson, Northwestern University, Evanston, Illinois. We thank the anonymous reviewers for their helpful suggestions during the preparation of this manuscript.

Brown, W. S., Rothman, H. B., and Sapienza, C. M. (**2000**). "Perceptual and acoustic study of professionally trained versus untrained voices," J. Voice **14**, 301–309.

Burnett, T. A., and Larson, C. R. (**2002**). "Early pitch-shift response is active in both steady and dynamic voice pitch control," J. Acoust. Soc. Am. **112**, 1058–1063.

Burnett, T. A., Senner, J. E., and Larson, C. R. (**1997**). "Voice F0 responses to pitch-shifted auditory feedback: A preliminary study," J. Voice **11**, 202–211.

Burnett, T. A., Freedland, M., Larson, C. R., and Hain, T. (**1998**). ''Voice F0 responses to manipulations in pitch feedback,'' J. Acoust. Soc. Am. **103**, 3153–3161.
Clarkson, J. K., and Deutsch, J. A. (**1966**). ''Effect of threshold reduction on the vibrato,'' J. Exp. Psychol. **71**, 206–210.
Dejonckere, P. H. (**1995**). ''Fascinating and Intriguing Vibrato,'' in *Vibrato*, edited by P. H. Dejonckere, M. Hirano, and J. Sundberg (Singular, London).
Deutsch, J. A., and Clarkson, J. K. (**1959**). ''Nature of the vibrato and the control loop in singing,'' Nature (London) **183**, 167–168.
Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). ''Effects of frequency-shifted auditory feedback on voice F0 contours in syllables,'' J. Acoust. Soc. Am. **111**, 357–366.
Hain, T. C., Burnett, T. A., Larson, C. R., and Kiran, S. (**2001**). ''Effects of delayed auditory feedback (DAF) on the pitch-shift reflex,'' J. Acoust. Soc. Am. **109**, 2146–2152.
Hain, T. C., Larson, C. R., Burnett, T. A., Kiran, S., and Singh, S. (**2000**). ''Instructing participants to make a voluntary response reveals the presence of two vocal responses to pitch-shifted stimuli,'' Exp. Brain Res. **130**, 133–141.
Horii, Y. (**1989**). ''Frequency modulation characteristics of sustained /a/ sung in vocal vibrato,'' J. Speech Hear. Res. **32**, 829–836.
Horii, Y., and Hata, K. (**1988**). ''A note on phase relationships between frequency and amplitude modulation in vocal vibrato,'' Folia Phoniatr. **40**, 303–311.
Hsaoi, 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**, 224–329.
Inbar, G., and Eden, G. (**1983**). ''Physiological evidence for central modulation of voice tremor,'' Biol. Cybern. **47**, 1–12.
Jones, J. A., and Munhall, K. G. (**2000**). ''Perceptual calibration of F0 production: evidence from feedback perturbation,'' J. Acoust. Soc. Am. **53**, 1246–1251.
Kawahara, H., and Williams, J. C. (**1996**). ''Effects of auditory feedback on voice pitch trajectories: characteristic responses to pitch perturbations,'' in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego).
King, J. B., and Horii, Y. (**1993**). ''Vocal matching of frequency modulation in synthesized vowels,'' J. Voice **7**, 151–159.
Koda, J., and Ludlow, C. (**1992**). ''An evaluation of laryngeal muscle activation in patients with vocal tremor,'' Otolaryngol.-Head Neck Surg. **107**, 684–696.
Lane, H., and Tranel, B. (**1971**). ''The Lombard sign and the role of hearing in speech,'' J. Speech Hear. Res. **14**, 677–709.
Larson, C. R., White, J. P., Freedland M. B., and Burnett, T. A. (**1996**). ''Interactions between voluntary modulations and pitch-shifted feedback signals: Implications for neutral control of voice pitch,'' in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego).
Laukkanen, A. M. (**1994**). ''Artificial pitch changing in auditory feedback as a possible method in voice training and therapy,'' Folia Phoniatr Logop **46**, 86–89.
Lippold, O. (**1971**). ''Physiological Tremor,'' Sci. Am. **224**, 65–71.
Marin, C. M., and McAdams, S. (**1991**). ''Segregation of concurrent sounds. II: Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width,'' J. Acoust. Soc. Am. **89**, 341–351.
Marsden, C. D. (**1984**). ''Origins of Normal and Pathological Tremor,'' in *Disorders in Movement: Tremor*, edited by L. J. Findley and R. Capildeo (Oxford U. P., New York).
Niimi, N., Horigushci, S., Kobayashi, N., and Yamada, M. (**1988**). ''Electromyographic study of vibrato and tremolo in singing,'' in *Vocal Physiology: Voice Production, Mechanisms, and Function*, edited by O. Fujimura (Raven, New York).
Orlikoff, R. F., and Baken, R. J. (**1989**). ''The effect of heartbeat on vocal fundamental frequency perturbation,'' J. Speech Hear. Res. **32**, 576–582.
Prame, E. (**1995**). ''Measurement of the vibrato rate of ten singers,'' in *Vibrato*, edited by P. H. Dejonckere, M. Hirano, and J. Sundberg (Singular, London).
Ramig, L., and Shipp, T. (**1987**). ''Comparative measures of vocal tremor and vocal vibrato,'' J. Voice **1**, 162–167.
Rothenberg, M., Miller, D., and Molitor, R. (**1988**). ''Aerodynamic investigation of sources of vibrato,'' Folia Phoniatr. **40**, 244–260.
Sapir, S., and Larson, K. K. (**1993**). ''Supralaryngeal muscle activity during sustained vibrato in four sopranos: Surface EMG findings,'' J. Voice **7**, 213–218.
Sapir, S., and McClean, M. D. (**1983**). ''Effects of frequency-modulated auditory tones on the voice fundamental frequency in humans,'' J. Acoust. Soc. Am. **73**, 1070–1073.
Seashore, C. E. (**1938**). *Psychology of Music* (McGraw–Hill, New York).
Seidner, S., Nawka T., and Cebulla, M. (**1995**). ''Dependence of the vibrato on pitch, musical intensity, and vowel in different voice classes,'' in *Vibrato*, edited by P. H. Dejonckere, M. Hirano, and J. Sundberg (Singular, London).
Shipp, T., and Izdebski, K. (**1981**). ''Current evidence for existence of laryngeal macrotremor and microtremor,'' J. Forensic Sci. **26**, 501–505.
Shipp, T., Sundberg, J., and Haglund, S. (**1984**). ''A model of frequency vibrato,'' in *Transcripts of the 11th Symposium Care of the Professional Voice*, edited by L. van Lawrence (Voice Foundation, New York), pp. 116–117.
Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois U.P., Dekalb), pp. 168–170.
Sundberg, J. (**1995**). ''Acoustic and psychoacoustic aspects of vocal vibrato,'' in *Vibrato*, edited by P. H. Dejonckere, M. Hirano, and J. Sundberg (Singular, London).
Titze, I. R. (**1996**). ''Coupling of neural and mechanical oscillators in control of pitch, vibrato, and tremor,'' in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego).
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**, 215–223.
Titze, I. R., Story, B., Smith, M., and Long, R. (**2002**). ''A reflex resonance model of vocal vibrato,'' J. Acoust. Soc. Am. **111**, 2272–2282.
Vennard, W. (**1967**). *Singing: The Mechanism and the Technique* (Carl Fisher, New York).
Winckel, F. (**1974**). ''Acoustic cues in the voice for detecting laryngeal diseases and individual behavior,'' in *Ventilatory and Phonatory Control Systems: An International Symposium*, edited by B. Wyke (Oxford U. P., London).
Xu, Y., Larson, C. R., and Bauer, J. J. (**2002**). ''On-line processing of voice pitch feedback during production of mandarin tones,'' presented at the Third Biennial International Conference on Voice Physiology and Biomechanics, Denver, CO.

Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study

Pascal Perrier[a)]
Institut de la Communication Parlée, UMR CNRS 5009, INPG, Grenoble, France

Yohan Payan
Laboratoire TIMC, CNRS, Université Joseph Fourier, Grenoble, France

Majid Zandipour and Joseph Perkell
Speech Communication Group, R.L.E., Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received 11 October 2002; revised 8 May 2003; accepted 9 May 2003)

This study explores the following hypothesis: forward looping movements of the tongue that are observed in VCV sequences are due partly to the anatomical arrangement of the tongue muscles, how they are used to produce a velar closure, and how the tongue interacts with the palate during consonantal closure. The study uses an anatomically based two-dimensional biomechanical tongue model. Tissue elastic properties are accounted for in finite-element modeling, and movement is controlled by constant-rate control parameter shifts. Tongue raising and lowering movements are produced by the model mainly with the combined actions of the genioglossus, styloglossus, and hyoglossus. Simulations of V1CV2 movements were made, where C is a velar consonant and V is [a], [i], or [u]. Both vowels and consonants are specified in terms of targets, but for the consonant the target is virtual, and cannot be reached because it is beyond the surface of the palate. If V1 is the vowel [a] or [u], the resulting trajectory describes a movement that begins to loop forward before consonant closure and continues to slide along the palate during the closure. This pattern is very stable when moderate changes are made to the specification of the target consonant location and agrees with data published in the literature. If V1 is the vowel [i], looping patterns are also observed, but their orientation was quite sensitive to small changes in the location of the consonant target. These findings also agree with patterns of variability observed in measurements from human speakers, but they contradict data published by Houde [Ph.D. dissertation (1967)]. These observations support the idea that the biomechanical properties of the tongue could be the main factor responsible for the forward loops when V1 is a back vowel, regardless of whether V2 is a back vowel or a front vowel. In the [i] context it seems that additional factors have to be taken into consideration in order to explain the observations made on some speakers. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1587737]

PACS numbers: 43.70.Bk, 43.70.Aj [AL]

I. INTRODUCTION

Many studies aimed at understanding the control strategies of speech production have been based on analyses of observable (articulatory or acoustic) speech signals. However, it is well known that comparable observed patterns could be produced by different underlying mechanisms.

For example, Nelson (1983) suggested that speech articulatory movements would be produced with an optimized control strategy aiming at minimizing the jerk (the third derivative of displacement versus time). At the same time, he demonstrated that the velocity profile associated with jerk minimization is bell-shaped and quite similar to the velocity profile of a second-order system. Since then, the kinematic properties of speech articulators have been shown to be close to those of a second-order dynamical model (see, for example, Ostry and Munhall, 1985, for tongue movements). The central question is, thus, to know whether these kinematic properties are the result of optimized central control (as implied by Nelson), whether they are a natural consequence of the biomechanical properties of the speech articulators, or whether they are the result of the combination of both effects.

Another example of this nature can be found in the studies initiated by Adams *et al.* (1993). These authors observed that when speaking rate decreases, the number of local maxima observed in the velocity profiles of articulatory movements (so-called velocity peaks) would increase from one or two to several. They suggested that the change from fast to slow movements would imply a drastic modification of the underlying control strategy from a single movement to a sequence of multiple submovements. However, a study carried out by McClean and Clay (1995) showed that the variability in the number of velocity peaks observed for an articulatory gesture across speaking rates could be related to the firing rate of motor units, which would naturally vary when velocity changes. Thus, far from being evidence of a drastic modification of the control, multiple velocity peaks could simply originate in the natural variation of a low-level neurophysiological process. Again, these observations raise

[a)]Electronic mail: perrier@icp.inpg.fr

the question of the relative influences of, on the one hand, control strategies and, on the other hand, physical, physiological, and neurophysiological properties, on the kinematic patterns observed during speech production.

In this context, the present paper proposes an assessment of the potential contribution of the biomechanics to complex articulatory patterns (called ''articulatory loops'') observed during the production of VCV sequences, where C is a velar stop consonant (Houde, 1967; Mooshammer *et al.*, 1995; Löfqvist and Gracco, 2002). The study is based on simulations made with a 2D biomechanical tongue model. After a summary of the main experimental observations of articulatory loops in the literature and their possible explanations, the tongue model will be presented and results of various simulations will be described that contribute to the analysis.

II. BACKGROUND: EXPERIMENTAL EVIDENCE OF "ARTICULATORY LOOPS"

Articulatory looping patterns were first described in 1967 by Houde, who analyzed cineradiographic data in a number of V1–[g]–V2 sequences, and noted that ''*a distinct forward directed gesture takes place during the closure*'' of the consonant. Studying tongue body motions from the trajectories of four radio-opaque markers attached to the midline of the tongue of a single speaker, Houde noted: ''*When the closure occurs during a forward directed vowel transition (/ugi/, /agi/), ... the contact appears to be sustained while sliding along the palate for a distance of up to 6 mm.*'' (Houde, 1967, p. 129). In these sequences, the observed sliding movement could easily be interpreted as the consequence of the vowel-to-vowel gesture (oriented from the back to the front). However, such a hypothesis would not be consistent with the other set of observations provided by Houde: ''*When the palatal closure occurs during a rearward movement of the tongue* [...]*, in some cases (/i'gagi/) its direction is temporarily reversed. It behaves as if forward movement had been superimposed, during contact, on the main rearward movement of the tongue.*'' (Houde, 1967, p. 129). In addition, similar movements patterns were also observed in V1–g–V2 sequences where V1=V2, thus apparently precluding an explanation based only on vowel-to-vowel coarticulation phenomena.

Houde suggested that the forward movement could result from a passive effect of forces generated on the tongue surface by the air pressure behind the contact location. Due to the closure of the vocal tract, the air pressure increases in the back cavity and could push the tongue in the forward direction. ''*The direction of the movement during closure is consistent with an increase in oral pressure, and as in the case of labial closures, a compliant element is required in the oral cavity, during the voiced palatal stop in order to sustain voicing. The passive reaction of the tongue may provide that required compliance.*'' (Houde, 1967, p. 133).

Since then, many additional observations have been made of such loops (Kent and Moll, 1972; Perkell *et al.*, 1993; Löfqvist and Gracco, 1994; Mooshammer *et al.*, 1995; Löfqvist and Gracco, 2002), and the hypothetical influence of air pressure in the back cavity has been analyzed further. Ohala has suggested that this looping movement could be ''*a very marked form of active cavity enlargement and could more than compensate for the other factors which disfavor voicing on velars.*'' (Ohala, 1983, p. 200). However, the hypothesis of active control of the loops has been seriously questioned by data collected on German speakers by Mooshammer *et al.* (1995). Their subjects produced articulatory loops during the unvoiced stop consonant [k] in [aka] that were even larger than for the voiced consonant [g]. This result clearly does not favor Ohala's hypothesis, but it does not refute the assumption that aerodynamics pressure forces could contribute to the forward movement observed in articulatory loops. Hoole *et al.* (1998) tried to assess the potential effect of the pressure forces quantitatively, by experimentally comparing the production of velar consonants during normal versus ingressive speech. Their results revealed forward articulatory loops in both conditions, but their size was significantly reduced in ingressive speech. This result supports the idea that aerodynamics could influence tongue movements, but, at the same time, it also indicates that other factors, perhaps including biomechanical properties of the tongue, may also contribute to generate the observed loops.

On the other hand, Löfqvist and Gracco (2002), inspired by studies of arm control movement in reaching or pointing tasks, suggested that the curvature of the articulatory trajectories that is at the origin of the looping patterns could arise from general motor control principles based on a cost minimization. Such a minimization would mean that the whole trajectory of the tongue would be planned, and that physical factors such as aerodynamics and biomechanics would play no direct role or a minimal role in the trajectory shape.

In this paper, we will explore a totally different hypothesis. Using Payan and Perrier's (1997) tongue model, we will assess the hypothesis that biomechanical factors may be at least partly responsible for the observed looping patterns.

III. THE TONGUE MODEL

Before giving details about the structure of the model, a short description of tongue *e* anatomy is provided, together with a brief overview of the state of the art in the field of the biomechanical modeling of the tongue.

A. A brief description of tongue structure

1. Tongue musculature

A detailed anatomical study of the tongue musculature has been described in Takemoto (2001). Thus, the description given here will only address functional aspects (Perkins and Kent, 1986) that were useful in the design of the 2D biomechanical tongue model. It will be limited to muscles for which the main influence can be described in the midsagittal plane, and muscles with fibers oriented mainly in the direction orthogonal to the midsagittal plane will be not presented. Most of the considered muscles are paired, with one on each side of the midsagittal plane; however, in the following description, their names are given in singular form. Among the ten muscles that act on the tongue structure, there are three extrinsic muscles that originate on bonystructures and insert into the tongue: the *genioglossus*, the *styloglossus*,

and the *hyoglossus*. They are responsible for the main displacement and shaping of the overall tongue structure (e.g., see Perkell, 1996). Contraction of the posterior fibers of the *genioglossus* produces a forward and upward movement of the tongue body, while its anterior fibers pull the anterior portion of the tongue downward. The *styloglossus* raises and retracts the tongue, causing a bunching of the dorsum in the velar region. The *hyoglossus* retracts and lowers the tongue body. Three additional intrinsic muscles, totally embedded in the tongue structure, contribute to a lesser extent to the sagittal tongue shape. The *superior longitudinalis* muscle shortens the tongue, and bends its blade upwards. The *inferior longitudinalis* muscle depresses the tip. The *verticalis* fibers depress the tongue and flatten its surface.

2. Tongue innervation

Tongue innervation (carrying its motor supply and its sensory and proprioceptive feedback) doesn't involve the same kind of neural circuitry as does the control of human limb movements. Whereas human limb muscles are innervated by spinal nerves, the muscles of the vocal tract are innervated by cranial nerves, which have their nuclei in the brain stem. However, most of the principles governing limb motor control also apply to the control of tongue movements. For example, as for the limbs, the efferent commands that are conveyed to tongue muscles (by the hypoglossal nerve) are likely to be modulated by proprioceptive feedback. Indeed, most of the oral mucosa, and particularly the tongue surface, is supplied with several different types of mechanoreceptors, and muscles spindles have been found within the tongue musculature (Cooper, 1953; Walker and Rajagopal, 1959; Porter, 1966; Trulsson and Essick, 1997).

B. The 2D biomechanical model of the tongue

A number of biomechanical tongue models have been developed already to study speech production (for 2D models, see Perkell, 1974, 1996; Kiritani *et al.*, 1976; Hashimoto and Suga, 1986; Honda, 1996; Sanguineti *et al.*, 1997; Payan and Perrier, 1997; Dang and Honda, 1998; for 3D models, see Wilhelms-Tricarico, 1995; Kakita *et al.*, 1985). The tongue model used in the current study represents a significant improvement of Payan and Perrier's 2D tongue model (1997). In this section, the fundamental aspects of the modeling approach are described in detail.

1. Biomechanical structure

An important first choice in modeling tongue structure was to limit the complexity of the model by only representing tongue characteristics that are relevant for speech. For this reason only the muscles mainly active during speech have been incorporated (see below for details). In addition, the tongue description has been limited to the midsagittal plane, in accordance with phonetic classifications of speech sounds that are based either on the position of the highest point of the tongue in this plane (Straka, 1965), or on the position of the vocal-tract constriction along the midline going from the glottis to the lips (Fant, 1960; Wood, 1979). This choice is also consistent with the fact that the kinematic data available in the literature describe tongue movements in the midsagittal plane (cineradiographic and electromagnetic recordings).

In order to develop a biomechanical model as close as possible to the morphological and physical characteristics of a given speaker, a native speaker of French, PB, who has already been the subject for a large number of acoustic and articulatory data recordings (by cineradiography, electropalatography, electromagnetography, and MRI), was employed as our reference speaker.

The PB vocal-tract contours (hard palate, velar region, pharynx, and larynx), shape and position of the mandible, the lips and hyoid bone, and the surface contour of the tongue were extracted from a lateral x-ray image of PB's head during a pause in a speech utterance. The corresponding tongue shape is, therefore, considered to be close to the rest position of the articulator.

The tongue is composed of a rich mixture of muscle fibers, glands, connective and fatty tissues, blood vessels, and mucosa. However, for a first approximation, only two categories of tissues were modeled: the *passive tissues* and the *active tissues*. The first category includes the mucosa, the connective and fatty tissues, blood vessels, and glands; the second category corresponds to muscle fibers. Measurements can be found in the literature showing that the stiffness of muscular tissues increases with muscle activation (Duck, 1990; Ohayon *et al.*, 1999). This feature is included in the model, by increasing the stiffness of the elements associated with an active muscle.

The *finite-element method* (FEM) was chosen to discretize the partial differential equations that describe continuous tongue deformations. These equations were established according to the laws of the linear elasticity. In addition to a precise description of the continuous, viscoelastic, and incompressible properties of a body, the FE method makes it possible, via the notion of *element*, to attribute specific biomechanical properties to individual regions of the structure. This feature is crucial in order to make a distinction between passive and active tissues that constitute the model.

Defining and distributing the elements inside the structure (i.e., the *mesh* definition) was the next stage of the procedure. This is a critical phase that involves a compromise among faithfulness to reality, design complexity, and computation time. Automatic mesh generators might have been used here but were not, mainly because we wanted the finite-element model of the tongue to represent its muscular anatomy. For this reason, the element geometries were designed manually, with specific constraints in term of (1) the number of elements and (2) anatomical arrangement of the main muscular components. Ideally, it would be optimal to design an FEM structure where the limits of the tissues and muscles could be mapped exactly into the geometry of the different elements.

Figure 1 presents the results of the manual FE mesh design: 221 nodes (intersections of lines in the figure) define 192 quadratic elements (areas enclosed by lines) located in-

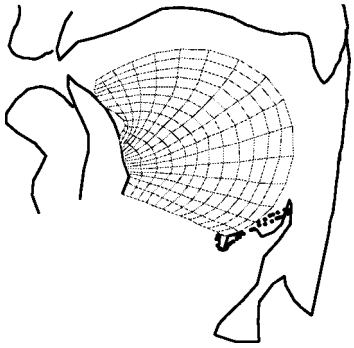


FIG. 1. Mesh of the 2D finite element tongue model in its rest position. The external vocal tract contours were extracted from x-ray data collected on the reference speaker PB.

side the sagittal tongue contour. Tongue attachments to the jaw and hyoid bone were modeled by allowing no displacement of the corresponding nodes, while tongue base support (essentially the effect of the mylohyoid muscle) was modeled by a reaction force, which prohibits any downward movements of the nodes located between the genioid tubercle of the mandible and the hyoid bone.

2. Muscle model

This section describes the definition of the insertion points and fiber orientations for the selected muscles, and the model of force generation that was adopted.

Asymmetries of tongue shapes in the lateral direction have been found in many experimental studies (see, for example, Stone *et al.*, 1997). However, to our knowledge, it has not been suggested that asymmetries would result from an explicit control. Consequently, our approach models the two symmetrical parts of each tongue muscle pair as a single entity. Only action in the midsagittal plane is considered.

a. Insertions and directions of the muscle fibers. Muscles are represented in the model at two different levels (see Fig. 2). First, their action on the tongue body is accounted for by ''macrofibers'' that specify the direction of the forces and the nodes of the FE mesh to which the forces are applied. Macrofibers are muscle-specific aggregations of segments (the bold lines in Fig. 2) connecting a number of selected nodes of the FE mesh to one another and to points on the bony structures (hyoid bone, jaw, styloid process). As depicted in Fig. 2, each muscle is composed of one to seven macrofibers, over which the global muscle force is distributed. Muscle force generation is modeled in a functional way according to Feldman's equilibrium point hypothesis of motor control (Feldman, 1986). More details will be given below.

Second, since the recruitment of a muscle modifies the elastic properties of the muscle tissues, muscles are also represented in the model by a number of selected elements within the FE structure (gray shaded elements in Fig. 2), whose mechanical stiffness increases with muscle activation. Since the model is limited to a 2D geometrical representation of the tongue, the association between elements and muscles, depicted in Fig. 2, was made on the basis of a simplified projection of the tongue in midsagittal plane. Special attention was devoted to assuring that the definition of the macrofibers, the geometry of the elements, and their assignment to muscles preserve the main properties of tongue muscle anatomy (Netter, 1989; Takemoto, 2001). Note that, because the inferior longitudinalis is a thin muscle, it was represented as a single macrofiber running from the hyoid bone to the tongue tip.

b. Modeling the generation of muscle force. To model the generation of muscle force, Feldman's ''λ model'' (Feldman, 1966, 1986), also referred to as the *equilibrium point hypothesis* (EPH), was used. This model, introduced for arm motor control, has been employed by Flanagan *et al.* (1990) and Laboissière *et al.* (1996) in their model of the jaw and hyoid bone complex. The λ model reflects the claim that α motoneuron (MN) activation, which generates force, is not centrally controlled, but is the consequence of the interaction between a central command and proprioceptive feedback. Feldman (1986) assumes that the central nervous system (CNS) independently acts on the membrane potentials of α and γ MNs in a way that establishes a threshold muscle length, λ, at which muscle activation starts. As the central command specifies changes in λ, muscle activation, and hence force, vary in relation to the difference between the actual muscle length and λ. Moreover, due to reflex damping, this activation also depends on the rate of change of muscle length. Feldman assumes that the nervous system regulates the equilibrium point of the muscle-load system by shifting the central command λ, in the form of changes in the central facilitation of MNs, producing a movement to a new equilibrium position.[1]

In the present model, consistent with the experimental force-length measurements reported by Feldman and Orlovsky (1972) for a cat gastrocnemius muscle, the relation between active muscle force and muscle activation is approximated by an exponential function (see Laboissière *et al.*, 1996; Payan and Perrier, 1997; Sanguineti *et al.*, 1997 for more details).

3. Elastic tissue properties

In the absence of any muscle recruitment, the tongue mesh represents passive tissues. Under these conditions, the model consists of an isotropic linear FE structure, whose biomechanical characteristics were chosen in order to model tissue quasi-incompressibility and to replicate mechanical measurements available in the literature.

Accounting for tissue incompressibility would require measuring tissue deformations in 3D space. This can obviously not be done properly in relation to a planar model. In this context, tongue deformations in the direction orthogonal to the midsagittal plane were assumed to be negligible in comparison to the geometrical changes in this plane (the so-called *plane strain hypothesis*). In this case, tissue quasi-incompressibility is equivalent to area conservation and can be modeled with a Poisson's ratio value close to 0.5 (Zienkiewicz and Taylor, 1989). This hypothesis is well supported by 3D measurements of tongue deformation during speech production, such as the ultrasound data published by Stone *et al.* (1997) or the MRI data analyzed by Badin *et al.* (2002).

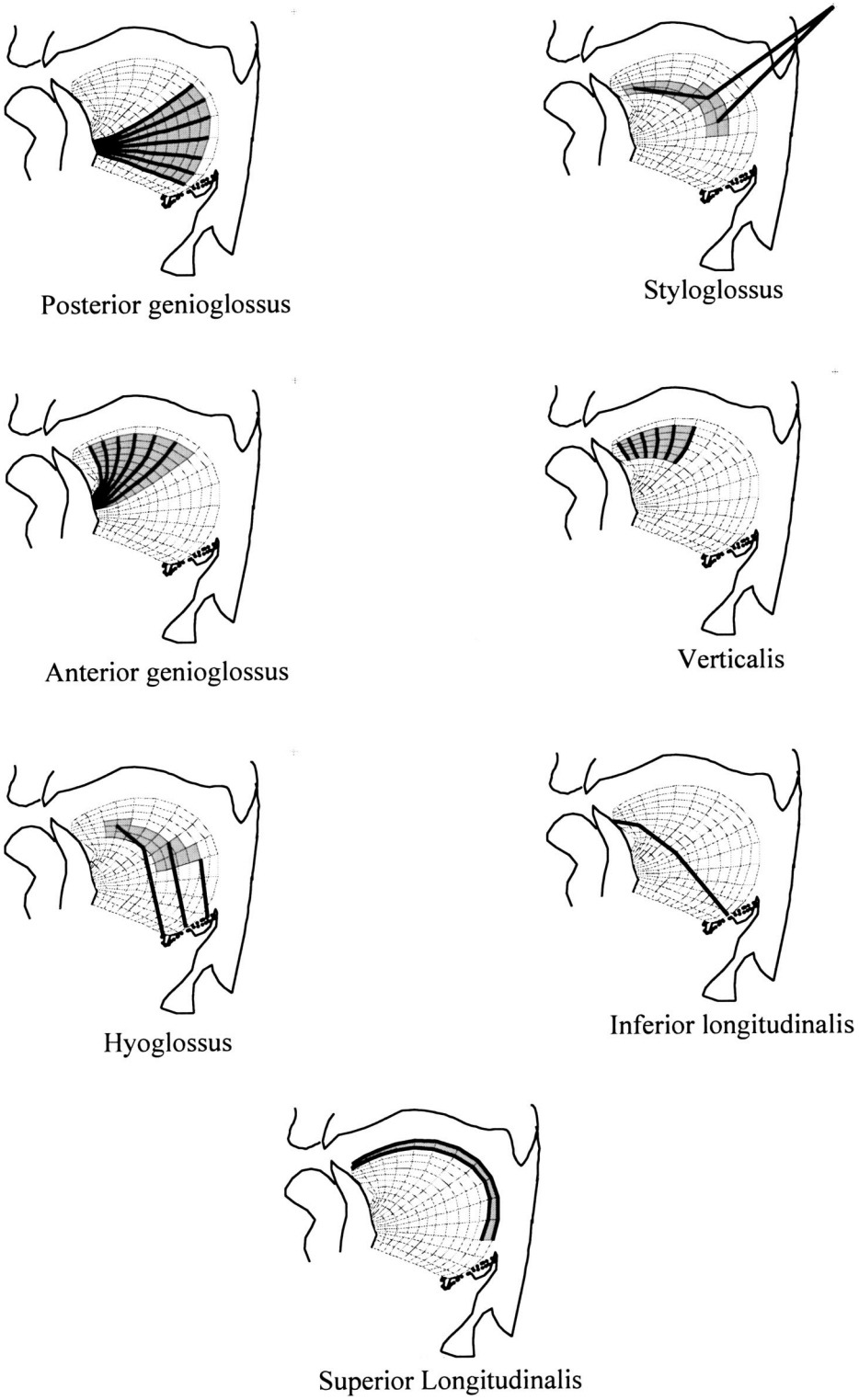


FIG. 2. Representation of the seven muscles taken into account in the model. The bold lines represent macrofibers, over which the global muscle force is distributed. The gray-shaded quadrilaterals are selected elements within the FE structure, whose mechanical stiffness increases with muscle activation.

The small-deformation framework of the FE method provides a method for stiffness modeling through the definition of the Young's modulus E value, which is assumed to fit the tissue stress–strain relationship (Zienkiewicz and Taylor, 1989). To our knowledge, no data are available in the literature about value of E value for passive tongue tissues, but measurements are reported for other part of the human body. Young's modulus values are estimated around 15 kPa for skin (Fung, 1993), 10 kPa for blood vessels, and between 10 and 30 kPa for vocal folds (Min *et al.*, 1994). To a first approximation, and after a number of trials, the value of Young's modulus of passive tongue tissues stiffness was set at $E_{\text{passive}} = 12.25$ kPa. With this value the temporal characteristics of tongue movements are realistic, as compared with data collected on real speakers, and the levels of force generated by the main muscles (GGp, GGa, STY, and HYO) are between 0.5 and 1.5 Newtons, which seems to be reasonable (Bunton and Weismer, 1994).

As mentioned earlier, when a muscle is activated, its fiber stiffness increases. Measured values for human skeletal

muscles have been reported to be 6.2 kPa for muscles at rest, and 110 kPa for the same muscles in a contracted position (Duck, 1990). The stiffness of cardiac muscle has been measured at close to 30 kPa at rest, and as high as 300 kPa when the muscle is activated (Ohayon *et al.*, 1999). In the framework of the FE method, modeling the global increase of muscle stiffness with activation was made possible by increasing the value of Young's modulus of muscular elements. Thus, the value of Young's modulus varied with muscle activation (between E_{passive} at rest, and $E_{\max}$ when the muscle is maximally contracted), while other tongue elements have a constant value of E equal to E_{passive}. For the present version of the model, and again after a number of trials, $E_{\max}$ was fixed at 100 kPa. Because of the various sizes of the muscles, this maximal value is reached for muscle dependent levels of force. Thus, for example, it is reached for a 2.8 N force for the posterior genioglossus and for a 0.8 N force for the hyoglossus. For a force level corresponding to normal speech conditions (i.e., between 0.5 and 1.5 N) the Young's modulus varies between 40 and 75 kPa.

Finally, these elastic parameters were validated by comparing the deformations of the FE structure induced by each tongue muscle with the deformations observed during real speech. The force developed by each muscle was tuned so that the global level of force produced during a tongue displacement was at a level similar to those measured on human tongue during reiterant speech production (Bunton and Weismer, 1994), and the direction and amplitude of the simulated deformations were verified to be compatible with data measured on PB during speech sequences (Badin *et al.*, 1995).

Figure 3 plots the tongue deformations induced by each modeled muscle. The tongue shapes shown in the figure are similar to those seen in a number of cineradiographic studies of speech movements (e.g., Perkell, 1969; Bothorel *et al.*, 1986). Note, however, that the upward curvature of the tongue generated by the action of the superior longitudinalis (lowest panel) is not sufficient when compared to real tongue tip deformations. Alternative implementations for this muscle, such as the one proposed by Takemoto (2001), are currently being tested.

4. Implementation of tongue–palate contacts in the biomechanical model

During the production of stop consonants, contacts between the tongue and palate dramatically influence tongue trajectories. Therefore, modeling collisions between the upper tongue contour and the palatal contour is necessary. In the present work, this includes two steps, which aim at: (1) detecting the existence of tongue/palate contact and (2) generating resulting contact forces.

From a theoretical point of view, solving the problem of the contact detection between a solid curved surface and deformable structure is quite complex. However, it is simplified considerably here for two reasons. First, the representation is two-dimensional instead of three-dimensional. Second, the contours delimiting the two bodies in contact (tongue against palate) are represented by points connected by straight lines. Under these simplified conditions, the contact detection problem is reduced to the detection of the intersection between two straight lines.

The force applied to the tongue when contact with the palate occurs was calculated according to a method originally proposed by Marhefka and Orin (1996). It is a so-called *penalty method*, based on a nonlinear relationship between the contact force and the positions and velocities of the nodes located on the upper tongue contour in the contact area (also, see Perkell, 1974, 1996). The basic principle of this method is as follows. If a node located on the upper tongue contour moves beyond the limits represented by the palate contours, a repulsion force F is generated in order to push this node back, up to the point where interpenetration is no longer detected. This force, applied to a node of the tongue model that is in contact with the palate, is computed according to Eq. (1)

$$\begin{aligned} &\mathbf{F} = (-\alpha \cdot x^n - \mu \cdot \dot{x} \cdot x^n) \cdot \mathbf{k} \quad \text{if } x < 0 \\ &\mathbf{F} = 0 \quad \text{if } x \geqslant 0, \end{aligned} \tag{1}$$

where x is the interpenetration distance (always a negative value when contact exists) between the node on the dorsal tongue contour and its orthogonal projection onto the palate contour; $\dot{x}$ is the first time derivative of the interpenetration distance; α is a coefficient representing the ''stiffness'' of the collision (a large α corresponds to hard contact); μ represents the ''damping factor'' of the collision; n accounts for the nonlinearity; and $\mathbf{k}$ is a unit vector orthogonal to the palate contour.

As emphasized in Eq. (1), the *penalty method* first tolerates a slight penetration of the tongue into the palate; then, it generates a force that pushes the node outward until the interpenetration distance is positive, at which point the contact force vanishes and the node is free to move back toward the palate. The cyclical behavior inherent to modeling contact in this way has a tendency to result in instabilities and oscillations. The parameters α, μ, and n have been arbitrarily fixed at *ad hoc* values ($\alpha = 60$; $\mu = 0.5$; $n = 0.8$), in such a way that, in the VCV simulations, the interpenetration distance and the amplitude of the potential oscillations remain smaller than a tenth of millimeter.

During contact, the tongue is free to slide along the palate. To our knowledge the viscosity constraining this sliding movement has never been measured. However, since the palate and the tongue are covered with saliva, and since saliva is a fluid that has lubricating properties, it is reasonable to assume that this viscosity factor is negligible as compared to the other damping factors that constrain tongue movement. Consequently, in the current model, no viscosity coefficient is used in the direction that is parallel to the palatal contour.

IV. SIMULATIONS WITH THE TONGUE MODEL

This section reports the results of a number of simulations that explore the potential role of biomechanical factors in the production of the looping articulatory patterns. The control of the tongue model is based on the concept that there is a separate target for each of the individual sounds of the sequence. Hence, specific target tongue shapes were first designed both for vowels and consonants, on the basis of

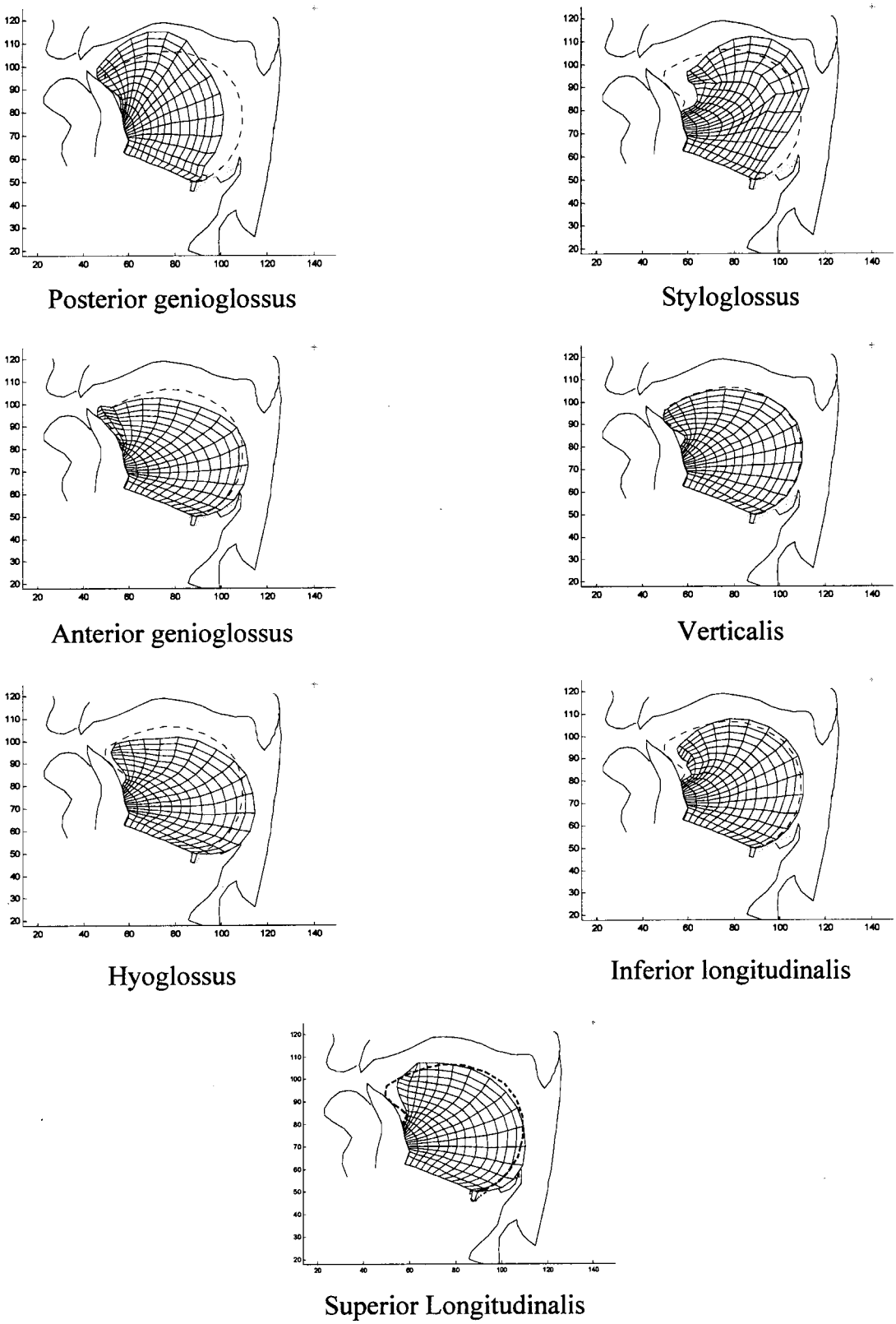


FIG. 3. Tongue deformations associated with muscle activations. The dotted line represents the contour of the tongue at rest. Units of *X* and *Y* axes are in mm.

data published in the literature for these sounds in similar contexts. Then, an initial set of simulations was generated for [aka], [uku], [iki], and [ika] sequences. In a second set of simulations, the effect on the articulatory trajectories of changes in the consonant target was studied. Finally, the effect of tongue palate interaction was analyzed specifically.

A. Underlying control of the tongue model during VCV sequences

1. Target-oriented control for vowels and consonants

As explained in Sec. III, muscle activations result from interactions between descending central control, specified by

the variables λ_i (index *i* referring to the different muscles), and the actual muscle lengths. A set of commands, λ_i, specifies the position of the tongue at which a stable mechanical equilibrium, also called posture, is reached. Feldman's suggestion is that arm movements are produced from posture to posture. In line with this hypothesis, in the current model, a sequence of discrete control variable values (λ_i), specifying successive postures, underlies a continuous trajectory through a sequence of phoneme targets. Movements are produced with constant rate shifts of the control variables from the settings of one target to those of the next target.

The phoneme targets represent the ideal goals toward which the tongue moves successively during the articulation of the sequence. For a given phoneme, these goals can vary with the phonetic context, since we also assume that their specification is the result of a higher-level planning process that takes into account the sequence as a whole and integrates some optimization principles. The description of this planning process is not part of the present paper (see, however, Perrier *et al.*, 1996a and Perkell *et al.*, 2000 for related discussions).

It is important to note that it is assumed that the underlying articulatory control is similar for vowels and consonants. However, the relation between the target specification and the tongue position actually reached differs significantly between these two classes of speech sounds. The specified vowel targets are located ventral to the palate contours; consequently, the corresponding tongue shape can actually be produced if the dynamical and time parametrization of the movement is adjusted appropriately. On the other hand, specified consonant targets are located beyond the palate and can therefore never be reached: they are ''virtual'' targets. Consequently, the tongue position reached during the production of the consonant is different from the specified one. It is the result of the combined influences of the target command and of contact between tongue and palate (cf. Perkell *et al.*, 2000). The virtual target hypothesis has been suggested by Löfqvist and Gracco for labial (1997) and for lingual (2002) stops, and it is supported by a kinematic comparison of articulatory data collected on German speakers and simulations made with the Payan and Perrier tongue model (Fuchs *et al.*, 2001).

2. Sequencing of the commands

Since the current study is focused on the influence of tongue biomechanics on articulatory paths, temporal as well as vowel-to-vowel coarticulatory effects (Öhman, 1966; Fowler, 1980; Perkell and Matthies, 1992; Matthies *et al.*, 2001) were purposely eliminated by making the following simplifications.

(i) No account is given, at the level of the articulation, of the differences between voiced and unvoiced consonants (see Löfqvist and Gracco, 1994, for examples of such differences); consequently, a unique articulatory target was used to specify the velar consonant in each vowel context. We arbitrarily refer to this consonant with the phonetic symbol [k].

(ii) Symmetrical temporal patterns have been chosen for the movements toward and away from the consonant.

(iii) The times of onsets and offsets of the motor command shift are the same for all muscles.

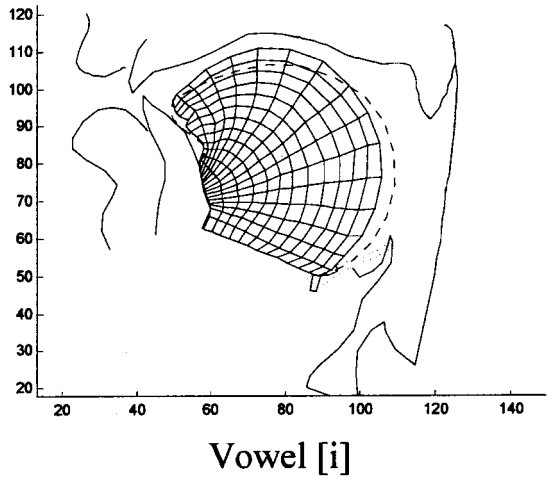
3. Selection of targets for [a], [i], [u], and [k]

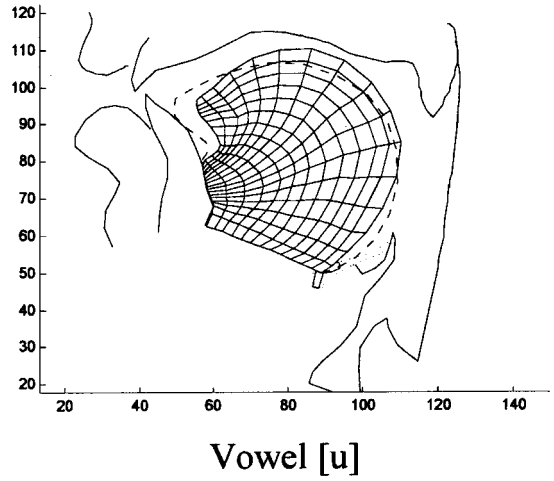
As discussed above, it is assumed that the target command for a phoneme results from a higher-level planning process that takes a set of successive targets into consideration. If this principle is applied strictly, the same phoneme, vowel or consonant, pronounced in two different phonetic sequences should be associated with two different target commands. For example, the target commands for [a] and [k] are likely to be different in the [aka] as compared to [aki] or [uka]. However, it is known that in V1CV2 sequences velar consonants are much more influenced by the surrounding vowels than the vowels are influenced by the consonant (Keating, 1993). Consequently, in order to minimize the number of simulations, in this work only the consonant target was assumed to vary as a function of the context. Thus, two different target commands were used for [k], a front one associated with front-vowel contexts, and a back one associated with back-vowel contexts, while a unique target was associated with each vowel.

These target commands were determined after a number of trials according to the following procedure. Knowing the main influence of each muscle on the tongue shape, we first approximated the muscle commands associated with each sound, by modifying them step by step, starting from the rest position, up to the point where a constriction was formed in the appropriate region in the vocal tract. Then, muscle commands were adjusted around this initial configuration so that tongue contours were reasonably close to data published in the literature for the same sound in similar vowel contexts (Houde, 1967, for English, and, for French, Bothorel *et al.*, 1986). To determine the two different muscle command sets for the velar consonant, special attention was given to the location and to the size of the contact region along the palate. Accordingly, for each new trial, simulations of [iki] and [aka] sequences were generated, and the shape of the tongue at different times during the consonantal closure was observed and compared to x-ray data. The evaluation criterion was to qualitatively replicate the differences in tongue shape observed experimentally for similar sequences. The resulting tongue shapes corresponding to the three vowels targets ([i],[a],[u]) are shown in Fig. 4, and the virtual targets associated with the two different muscle commands sets for the velar consonant, both in front and back context, are shown in Fig. 5.

The target defined for the vowel [i] involves activation of the posterior genioglossus (GGp) and, to a much lesser extent, of the styloglossus (STY). For [a], the target was obtained with recruitment of the hyoglossus (HG) and of the anterior genioglossus (GGa). The production of [u] is achieved with recruitment of the STY, and, to a much lesser extent, of the GGp.

For the velar consonant targets, three muscles are activated, namely the STY, the GGp, and, to a lesser extent, the inferior longitudinalis (IL). The balance between the forces





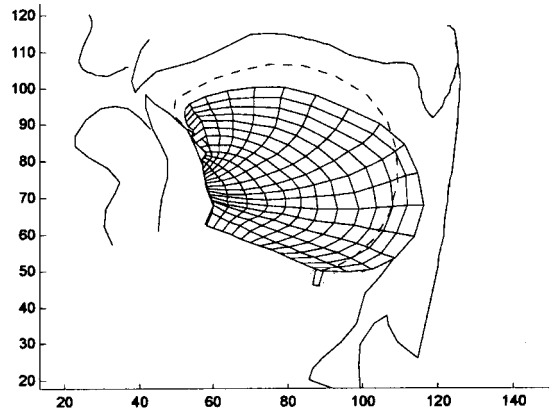
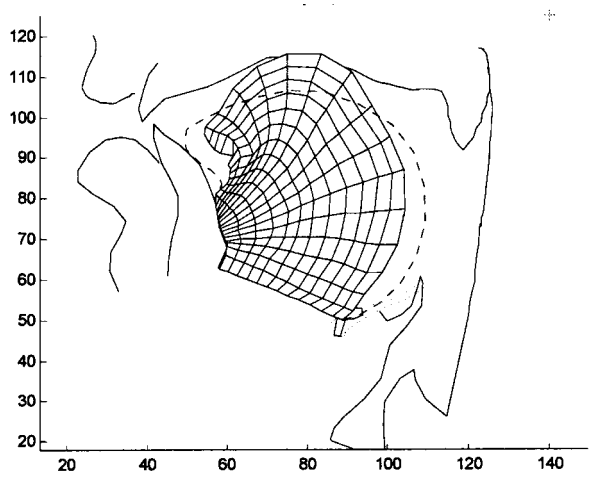
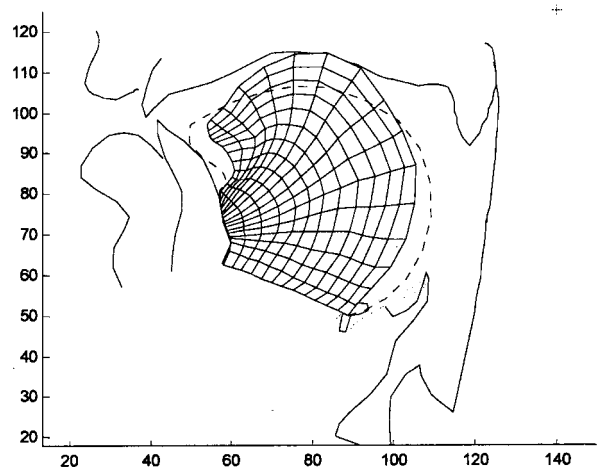


FIG. 4. Tongue shapes for the vowel targets used in the simulations. The dotted line represents the contour of the tongue at rest. Units of *X* and *Y* axes are in mm.





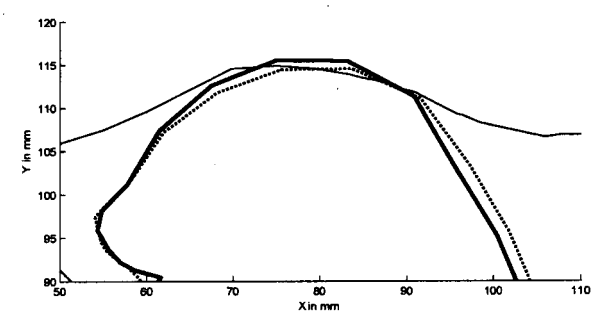


FIG. 5. Tongue shapes for the virtual consonant targets used in the simulations. Top panel: anterior target; middle panel: posterior target; in these two panels, the dotted line represents the contour of the tongue at rest; bottom panel: enlarged view of the tongue contours in the palatal region; dotted line: posterior target; solid bold line: anterior target. Units of *X* and *Y* axes are in mm.

produced by GGp and STY determines the difference between the anterior target and the posterior one. Figure 5 shows the corresponding overall tongue shapes (top and middle panels) and more closely in the palatal region (bottom panel). The highest point of the tongue is higher and more fronted for the anterior virtual target configuration. Under actual conditions, when the tongue is in contact with the palate, this difference induces for the anterior target a lengthening of the contact region towards the front, which is consistent with the observations provided for French stops by Bothorel *et al.* (1986) (pp. 180–181). In all cases the force due to gravity was not taken into account.

TABLE I. Timing of the commands for the VCV sequences.

	Duration (ms)
Vowel hold time	150
Vowel-to-[k] transition time	30
[k] hold time	100
[k]-to-vowel transition time	30
Vowel hold time	150

B. Simulation of V1–[k]–V2 sequences

1. Simulations for symmetrical V–[k]–V sequences

Simulations were first generated for [a], [u], and [i] in a symmetrical vowel context. The timing of the commands, the same for all the sequences, is given in Table I. For [a] and [u] the consonantal target was the posterior one, for [i] it was the anterior one. The trajectories of four nodes located in the palatal and velar regions on the upper contour of the tongue were analyzed for the three sequences (see Figs. 6, 7, and 8). For [aka] and [uku] we observe forward-looping patterns for the four nodes, with different amplitudes depending on the location of the nodes on the tongue and on the vowels: the loops observed in [a] context are clearly larger than in [u] context. For [iki] the movement is backward during the entire consonantal closure gesture; the size of the horizontal displacement is smaller than in the other two vowel contexts.

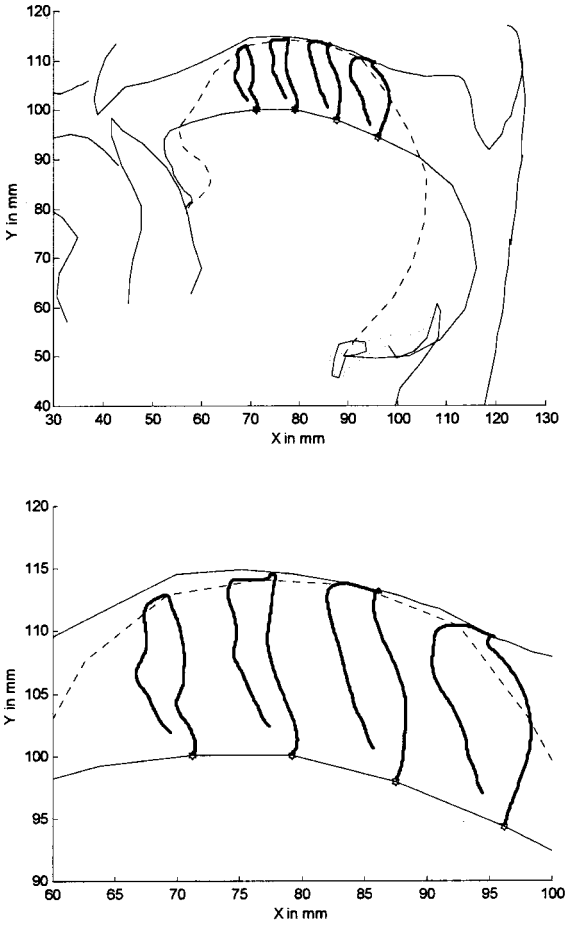


FIG. 6. Trajectories of four nodes on the dorsal contour of the tongue in the simulation of [aka]; top panel: general sagittal view; bottom panel: close up in the palatal region. The solid tongue contour represents the initial vowel configuration; the dotted line contour corresponds to the consonant configuration just before release; for each trajectory, the starting point is marked with a small open circle on the solid line tongue contour. Units of *X* and *Y* axes are in mm.

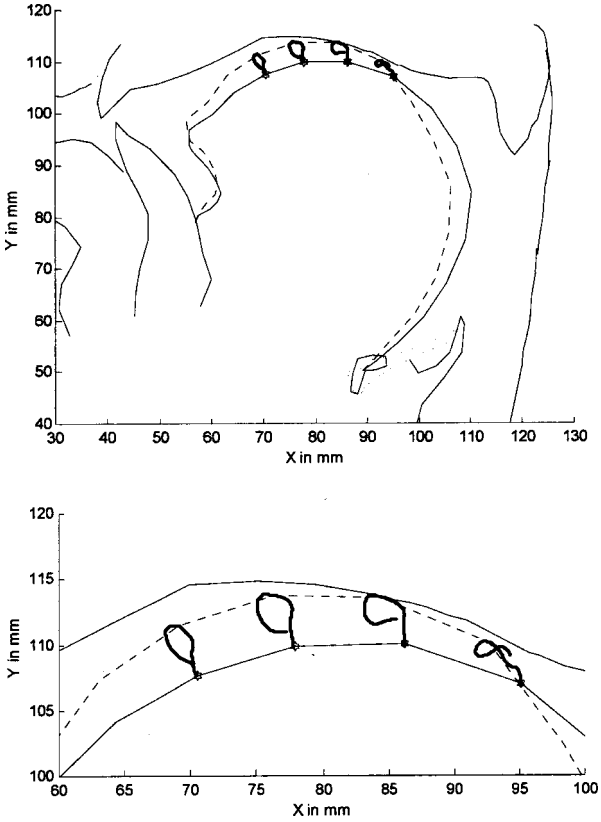


FIG. 7. Trajectories of four nodes on the dorsal contour of the tongue in the simulation of [uku]; top panel: general sagittal view; low panel: close up in the palatal region. The solid tongue contour represents the initial vowel configuration; the dotted line contour corresponds to the consonant configuration just before release; for each trajectory, the starting point is marked with a small open circle on the solid line tongue contour. Units of *X* and *Y* axes are in mm.

2. Simulations for asymmetrical V1–[k]–V2 sequences

Simulations were also made for asymmetrical sequences, where the vowels preceding and following the consonant were different. The timing of the commands was the same as in the symmetrical VCV simulations.

Special attention was devoted to the [ika] sequence, since Houde (1967) observed in some cases in [i'ga] a reversal of the main rearward movement during the consonant. As for [iki], the anterior target was used for the stop consonant.

Figure 9 shows the trajectories of the same four nodes. It can be observed that in this simulation, no reversal is produced and that the tongue slides continuously backward along the palate for about 2 mm during the [k] closure.[2] This result will be discussed later in Secs. IV C 1 and IV C 2 in relation to the consonant target location.

Concerning asymmetrical sequences V1–[k]–V2 in general, experimental studies (Houde, 1967; Mooshammer *et al.*, 1995, Löfqvist and Gracco, 2002) have systematically shown that sequences with V1=[i] show a much smaller amount of movement during the consonantal closure in comparison with V1=[u] or [a]. Figure 10 shows the trajectory

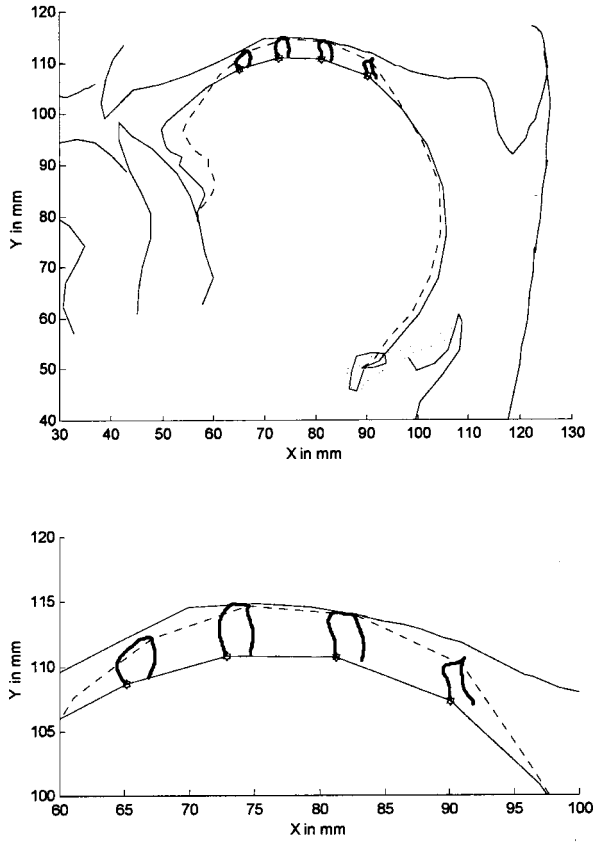


FIG. 8. Trajectories of four nodes on the dorsal contour of the tongue in the simulation of [iki]; top panel: general sagittal view; low panel: close up in the palatal region. The solid tongue contour represents the initial vowel configuration; the dotted line contour corresponds to the consonant configuration just before release; for each trajectory, the starting point is marked with a small open circle on the solid line tongue contour. Units of *X* and *Y* axes are in mm.

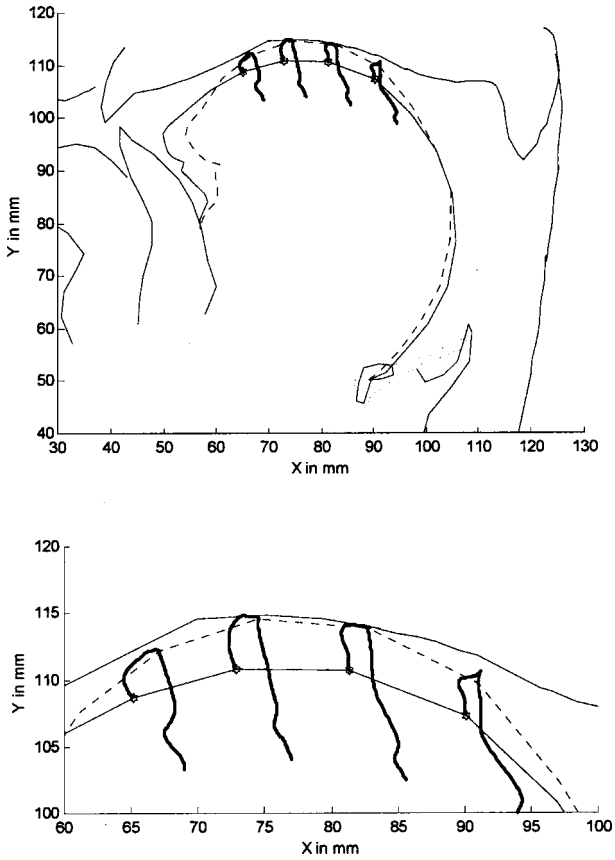


FIG. 9. Trajectories of four nodes on the dorsal contour of the tongue in the simulation of [ika]; top panel: general sagittal view; low panel: close up in the palatal region. The solid tongue contour represents the initial vowel configuration; the dotted line contour corresponds to the consonant configuration just before release; for each trajectory, the starting point is marked with a small open circle on the solid line tongue contour. Units of *X* and *Y* axes are in mm.

described by a node on the dorsal tongue contour for all the contexts. This node is the second from the back in Figs. 6–9. In order to see the influence of V1 on the amplitude of the sliding movement during the closure, the results obtained for the same V1 are grouped in the same panel. It can be observed that the size of the loop is determined by the first vowel V1, and that the general trend observed on the experimental data is replicated: if V1=[i] the amplitude of the movement is clearly smaller than in the other cases. However, the differences are not as large as the measurements provided by Mooshammer *et al.* (1995). These results will also be discussed below in relation to the consonant target location (see Sec. IV C 1 c).

C. Influence of target locations and tongue biomechanics

1. Analysis of the articulatory trajectories generated in the simulations

Three aspects of the articulatory trajectories warrant more in-depth analysis: (1) the direction (forwards or backwards); (2) the loop curvature and orientation (clockwise or counterclockwise); (3) the amplitude of the movement during the consonantal closure (the size of the loop). These properties will be analyzed separately, in relation to specific aspects of the model used to generate the sequences.

a. Direction of the paths. In summary, for [ak] and for [uk], the nodes located in the palatal region describe forward-oriented trajectories, while the movement is backward for sequences [ik] whether the following vowel is [a] or [i]. In the case of the vowels [u] and [a], the virtual target position of [k] is located anterior to the vowel targets (see Figs. 4 and 5, low panels). For vowel [i], the consonantal target is located slightly posterior to the [i] target (see Figs. 4 and 5, top panels). Therefore, it can be concluded that, in the model, the direction of the movement during the V–[k] sequences is defined by the locations of the vowel and the consonant targets relative to each other. This influence of the target locations could also be easily predicted with a simple kinematic model that would be controlled in a target-based manner. Consequently, the biomechanical properties of the tongue model do not seem to play any role in the determination of the main direction of the movement, i.e., whether it is forward or backward oriented. However, a kinematic model by itself would describe straight paths, and could not account for the fact that ''the horizontal and the vertical components of movement towards the target are pursued independently'' (Mooshammer *et al.*, 1995, p. 20). Both experimental data and our simulations show this phenomenon, since the trajec-

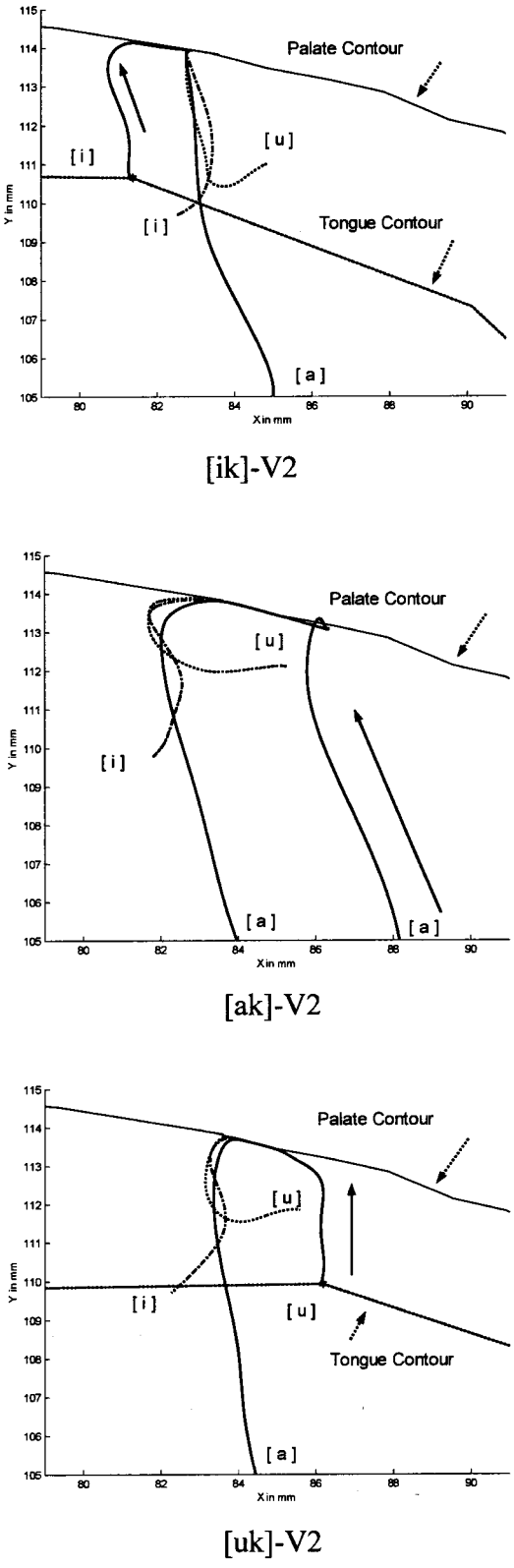


FIG. 10. Trajectories of a node on the dorsal tongue contour of the tongue (the second node from the back in Figs. 6–9) in the asymmetrical sequences; top panel: [ik]–V2 sequences; middle panel: [ak]–V2 sequences; bottom panel: [uk]–V2 sequences, where V2 is one of vowels [i] (dashed–dotted lines), [a] (solid lines), and [u] (dotted lines). The solid arrows show the directions of movements in the closing phase toward the consonant. Units of *X* and *Y* axes are in mm.

tories are curved. In our simulations the trajectory shapes are determined by the biomechanical properties of the model as explained below.

b. Trajectory curvature and loop orientation. The control underlying all the simulations presented above is extremely simple: the transition between two successive sounds is based on a linear interpolation between the two associated sets of muscle threshold lengths at the targets. Consequently, the curvature of the articulatory trajectories cannot be a direct consequence of the control itself. This phenomenon is due to the biomechanical properties of the tongue model, i.e., the passive tongue elasticity, the muscle arrangements within the tongue, and the force generation mechanism.

The passive elasticity is taken into account with the finite-element method. Thus, the continuous mechanics of tongue tissue is modeled: force acting on a specific part of the tongue has consequences on the whole tongue body. The relations among the strains generated in different parts of the tongue are nonlinear and depend on the finite-element parameters (Young's modulus and Poisson's ratio).

Muscle fiber orientations are not constant during a movement, since some of the muscle insertions are fixed (for example, the bony insertion of the styloglossus) while others are moving with tongue tissues (for example, the other ends of the styloglossus). As a consequence the directions of all muscle forces change during the movement.

Additional nonlinearities are introduced in the force generation due to the use of Feldman's control model. Because the model incorporates the concept of a threshold length, a muscle can suddenly become active if its length exceeds the threshold. Moreover, once a muscle is active, the force generated is an exponential function of its length. External forces are generated temporarily during the contact between tongue and palate, which adds another nonlinearity.

In our simulations, the combination of all these nonlinearities is responsible for the curved aspect of the trajectories. Thus, contrary to Löfqvist and Gracco's (2002) suggestion, it is not necessary to invoke a general optimization principle that would plan the entire trajectory to explain the trajectory shape.

The variation of the magnitudes and orientations of muscle forces during the movement, as determined by the combination of target commands which specify the time variation of the threshold muscle lengths, and tongue deformation, which modifies the length and the orientation of the muscle fibers, also contributes to the shape and orientation of the loop. For example, because of the combined actions of the GGp and the STY, for [aka] and [uku], the middle part of the tongue first moves upward and then forward before hitting the palate. It can also be observed that after the consonantal closure for [aka], [uku], and [iki], the first part of the movement toward the vowel is forward oriented, although only slightly so for [iki]. This forward movement is observed even if the subsequent vowel is posterior to the consonant release location, even though the motor commands do not specify movement in the forward direction. This result must therefore be a consequence of muscular anatomy and the tongue model's biomechanical properties.

c. Movement amplitude during the consonantal closure. It was noted in Sec. IV B, that, in all the V1–[k]–V2 simulations, the amplitude of the sliding movement of the tongue along the palate during the consonantal closure is mainly determined by the first vowel V1: the tongue slides over a distance of 5 mm for V1=[a], 3 mm for V1=[u], and around 2 mm for V1=[i].

In order to understand the origins of this phenomenon, different parameters were investigated: the amplitude and the orientation of the velocity vector just before consonant contact occurs, and the distances between nodes describing the tongue shape at the beginning of the consonantal closure and the virtual consonant tongue-shape target (the shape it would assume without interference from the palate). Additional simulations of the [aka] sequence were also calculated with various transition times from [a] to [k], in order to change the velocity while keeping the target commands constant.

Considering all these simulations, no clear relation could be found between, on the one hand, the magnitude and direction of the velocity vector just before the contact and, on the other hand, the amplitude of the movement during the closure. The only systematic finding is related to distance between the tongue shapes at the beginning of the closure and at the consonant virtual target. This is illustrated by Fig. 11, which shows these tongue shapes for [aka], [uku], and [iki] (from top to bottom). Considering the results depicted in Figs. 6–8, it can be seen that the length of the sliding contact section of the movement is related to the distance between the position of the tongue when it first contacts the palate (C) and the position of the consonant's virtual target (V). In the case of vowels [u] and [a], starting from the vowel, the tongue moves first upward and forward until it hits the palate. From this time, the vertical movement becomes strongly constrained by the palatal contour. Since the tongue shape at the first point of contact is posterior to the virtual target shape of the consonant, the tongue continues to slide forward along the palate in the direction of the virtual consonant target, and the larger the distance between the two shapes, the larger the amplitude of the sliding movement.

For vowel [i], the first part of the movement is upward and backward. The movement in the vertical direction becomes strongly constrained when the tongue hits the palate, slightly in front of the consonantal target (recall that, in this case, the anterior target was used). Consequently, the tongue slides along the palate in the backward direction over a small distance.

The virtual target for the consonant is specified at the control level. The tongue shape at the beginning of the consonantal closure is the result of the tongue deformation from the vowel, which depends on muscular anatomy and biomechanical properties of the tongue (see Sec. IV C 1 b) and on the virtual target specified for V1 and [k]. We have shown that the amplitude of the movement during the consonantal closure depends on the distance between these two tongue shapes and on the interaction with the palate. Consequently, the amplitude of the movement during the closure is the result of a combination of effects related to the control (the virtual target sequence) and to biomechanical factors.

These observations can also explain the differences between our simulations and Mooshammer *et al.*'s (1995) measurements about the size of the loops in various V1–[k]–V2: while the general orientation of the loop is the same for each speaker, the amplitude of the sliding movement during the closure depends on speaker specific properties, at a control and at a physical level.

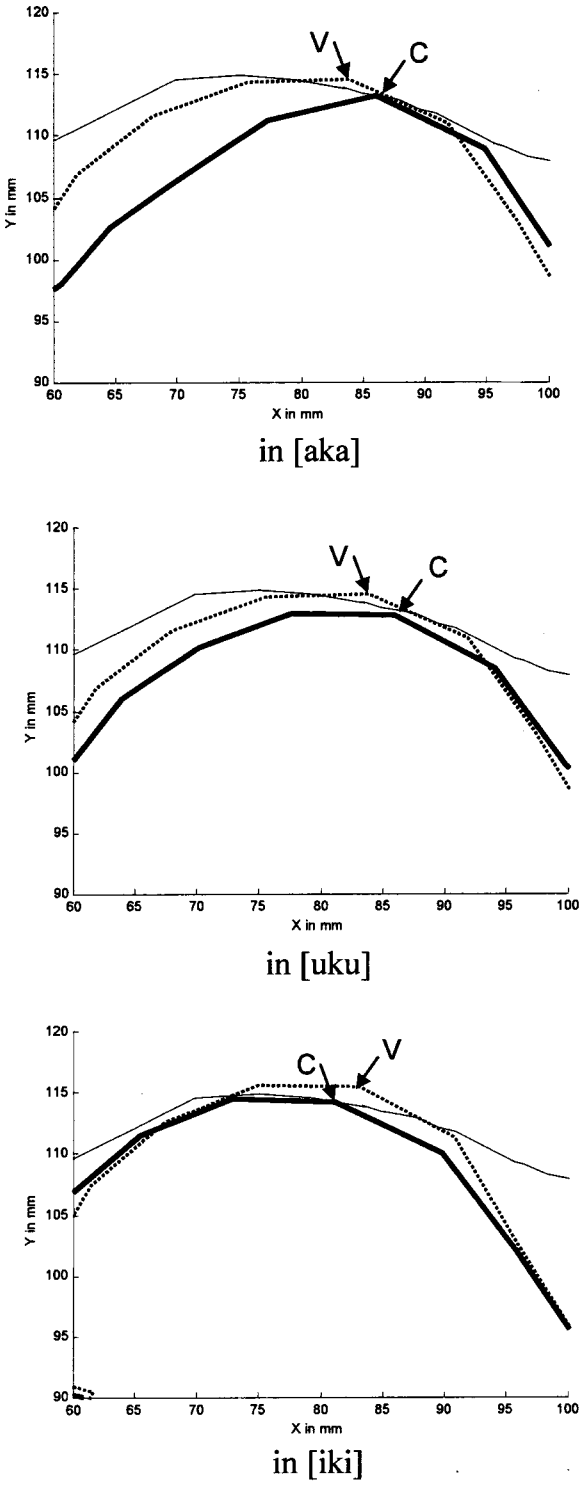


FIG. 11. Close up in the palatal region of the tongue shapes (bold line, labeled C) at the beginning of the consonantal closure for [aka], [uku], and [iki]. The dotted line represents the contour of the tongue at the virtual consonant target (labeled V). The solid line represents the palatal contour. Units of *X* and *Y* axes are in mm.

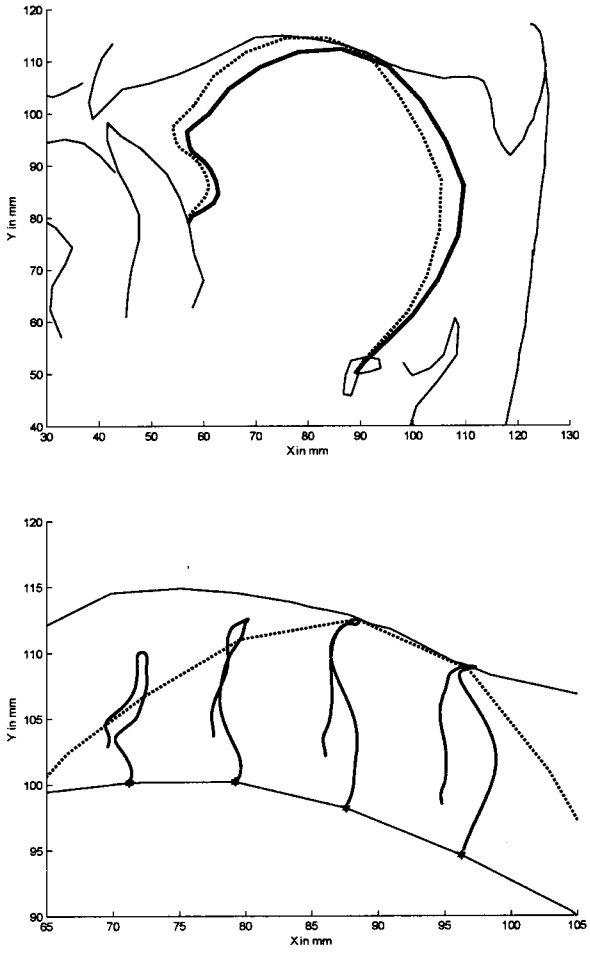


FIG. 12. Generation of backward-oriented loops for [aka] through target shifting. Top panel: original (dotted line) and modified (bold line) virtual consonant targets; lower panel: trajectories of four nodes on the dorsal contour of the tongue in the simulation of [aka] with the modified consonant target: close up in the palatal region. The solid tongue contour represents the initial vowel configuration; the dotted line contour corresponds to the consonant configuration just before release; for each trajectory, the starting point is marked with a small open circle on the solid line tongue contour. Units of *X* and *Y* axes are in mm.

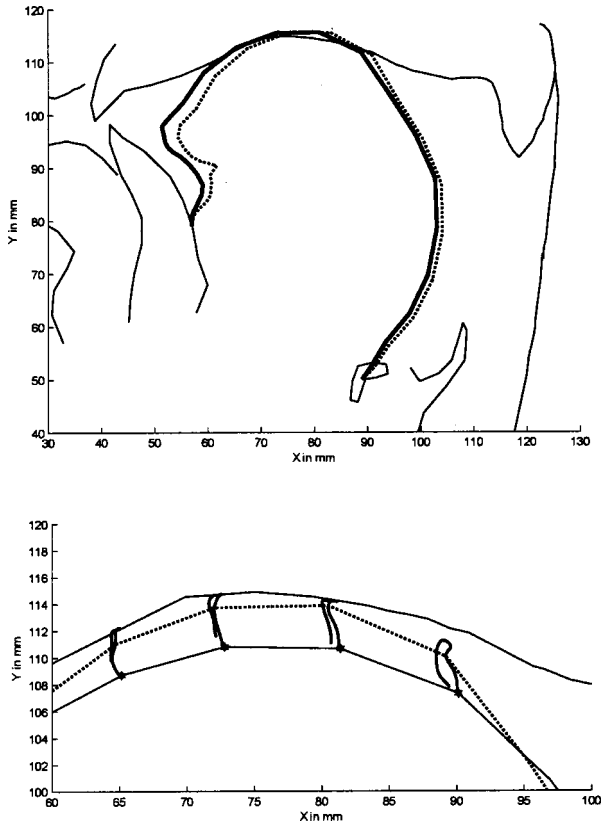


FIG. 13. Generation of forward-oriented loops for [iki] through target shifting. Top panel: original (dotted line) and modified (bold line) virtual consonant targets; lower panel: trajectories of four nodes on the dorsal contour of the tongue in the simulation of [iki] with the modified consonant target: close up in the palatal region. The solid tongue contour represents the initial vowel configuration; the dotted line contour corresponds to the consonant configuration just before release; for each trajectory, the starting point is marked with a small open circle on the solid line tongue contour. Units of *X* and *Y* axes are in mm.

2. Reversal of loop direction through consonant target shifting

We have seen that in our simulations the direction of the loops is determined by the positions of the consonant and vowel target tongue shapes relative to each other. In this context, it should be interesting to determine the extent to which the generated patterns are sensitive to changes in the specified locations of the targets. More specifically, we are interested in conditions that would cause the directions of the loops to be reversed. Hence, additional simulations were generated, where the consonant target was moved in the direction opposite to the originally observed loop direction: for [aka] and [uku], the target was gradually moved backward to a position determined by increasing the recruitment of STY and decreasing that of GGp; for [iki] it was moved forward by making the opposite changes in muscles recruitment.

Figure 12 shows the results of the simulation for [aka] in which the trajectory of the second node from the back becomes backward oriented. The top panel shows in dotted lines the virtual target used in [k] in the preceding simulations where forward-oriented loops were observed, and in solid line the virtual target obtained by modifying the recruitment of STY and GGP to produce a tongue contour that is positioned at the place where the first reversal of a node trajectory could be observed (see the bottom panel). The latter virtual target can be considered as a boundary within the vocal tract between two kinds of articulation for velar stops: for the virtual target tongue shapes that are more anterior than this boundary, the loops observed in [aka] will be forward oriented; for the virtual target tongue shapes that are more posterior, the loops will be backward oriented. Starting from the posterior target chosen for [k] in the preceding simulations, it took large changes in muscle commands to generate the differences in shape and to reverse the direction of the articulatory loop. As a consequence, the consonant target where reversal occurs is significantly different from the one used in the preceding simulations: the constriction is now essentially in the region of the soft palate, and not in the velo-palatal region as usually observed for the velar stops /g/ and /k/ (Bothorel *et al.*, 1986). Similar results were found for [uku].

Figure 13 shows the result obtained for [iki], with a

presentation identical to Fig. 12. It can be seen that, contrary to [aka], a small forward shift of the consonant target, associated with very small changes in muscle commands, was enough to reverse the direction and the orientation of the loops, which are now forward directed and counterclockwise oriented (as opposed to Fig. 8, in which they are backward directed and clockwise oriented). The latter consonant target is still reasonable for a [k] articulated in a front-vowel context (Bothorel *et al.*, 1986).

These results suggest that the forward direction of the looping patterns observed in the [a] and [u] contexts is very stable in the face of moderate changes in the consonant target location, while loop variability is likely to be observed in the [i] context, in which small perturbations of the target positions can reverse the loop direction and its orientation.

D. Effect of tongue–palate interaction

It can be concluded from the preceding section that, according to our model, target locations and tongue muscle anatomy and biomechanics, together with the tongue–palate interaction, may explain the existence, the direction, the orientation, and the size of the loops. In this section, the effect of tongue–palate interaction will be discussed more specifically.

The influence of tongue–palate interaction on the articulatory trajectories can be illustrated quantitatively in our model by generating the same VCV sequences in a ''virtual'' vocal tract where the palate is removed. In this case, the consonant target can actually be reached, and the corresponding articulatory trajectories can be observed and compared to the simulation with the palate. The trajectories obtained for [aka] under the same conditions as above (Sec. IV B), but with and without the palate, are shown in Fig. 14. The top panel shows trajectories of four nodes on the dorsal contour of the tongue in the simulation of [aka] in a virtual vocal tract without palate. The palatal contour is shown as a reference with a solid line. The lower (solid) tongue contour represents the initial vowel configuration. The open symbols show the locations of the nodes at the following successive times: circles—when node 3 passes upward through the palatal contour (initial contact for the consonant when the palatal constraint is in effect); squares—when node 2 passes upward through the palatal contour; and triangles—just before node 3 passes downward through the palatal contour. The lower panel shows a superimposition of the trajectories simulated with (dashed line) and without (solid line) the palate. The same muscle commands were used for both simulations.

From the moment the tongue goes above the palate (circles in Fig. 14, top half), the trajectories of both nodes 2 and 3 are oriented backward. The backward movement is more pronounced for node 3, due primarily to the conservation of volume constraint and the elastic properties of the model. In addition, the front part of the tongue (node 1), moves upward slightly after the central part of the tongue (nodes 2 and 3) has started to move downward (portions of the trajectories between the squares and triangles). In other words, the different nodes finish their upward movement at different times.

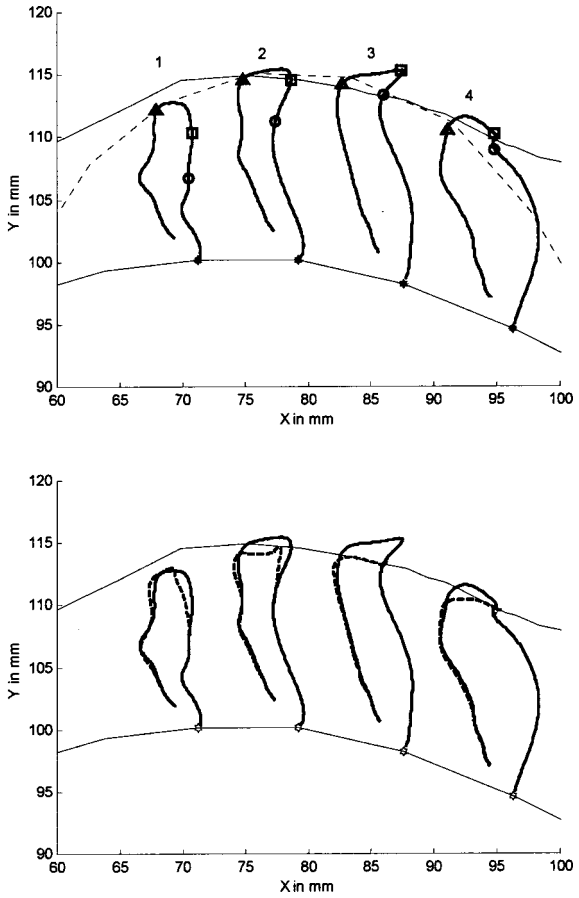


FIG. 14. Top panel: Trajectories of four nodes on the dorsal contour of the tongue in the simulation of [aka] in a ''virtual'' vocal tract without palate. The palatal contour is represented as a reference. The solid tongue contour represents the initial vowel configuration. The open symbols show the locations of the nodes at the following successive times: circles—when node 3 passes upward through the palatal contour (initial contact when the palatal constraint is in effect); squares—when node 2 passes upward through the palatal contour; and triangles—just before node 3 passes downward through the palatal contour (consonant release when the palatal constraint is in effect). The dotted contour corresponds to the virtual target of the consonant. The starting point is marked with a small filled circle on the solid line tongue contour. Lower panel: superimposition of the trajectories simulated with (dashed line) and without (solid line) palate. Units of *X* and *Y* axes are in mm.

The lower part of Fig. 14 shows that the four nodes initially follow the same trajectory in both simulations. However, as would be expected, differences appear when the tongue first reaches the palatal contour. In the absence of the palate, the tongue is free to continue its movement toward the virtual target without any limitation. From a little before the moment that the upper part of the tongue goes beyond the palatal line (circles in the top half of the figure), its movement is no longer continuously upward and forward. Especially for the two middle nodes, an upward and backward movement occurs first; then, the movement turns forward toward the virtual target location for the [k] (represented by the dotted line contour in the top panel). The backward movement is due to the fact that the force generated by the styloglossus becomes larger than the force generated by the posterior genioglossus. According to our model of muscle force generation (Sec. III B 2), force variation is due to changes in macrofiber lengths induced by tongue deformation. This particular influence of the styloglossus could not

be observed in simulations made with the palate, because, for the two middle nodes, the actions of the styloglossus and the genioglossus, combined with the reaction force generated by palatal contact, resulted in a force in the forward direction. In the simulations without the palate, there is no reaction force; therefore, the net force acting on theses nodes as the upper part of the tongue goes beyond the palatal contour is first oriented in the rearward direction, before again becoming forward oriented.

A comparison of the trajectories with the palate (dashed line, bottom half of Fig. 14) with those without the palate (solid line) shows that, after initial contact, the trajectories without the palate are slightly more posterior than the trajectories with the palate for the three anterior nodes. Thus, in the simulations, the interactions between the tongue and the palate influence the trajectory shape. The adequacy of such predictions could be tested in the future with actual articulatory behavior and the use of a device that measures the pressure of the tongue against the palate.

Note, however, that in both cases the VC portion of the trajectory is located well behind the CV portion, and that the maximum distance in the midsagittal plane between these two parts of the trajectory is not significantly modified by the presence or absence of palate. Therefore, in our model, the distance between the VC and the CV trajectories and the maximum size of the articulatory loops in the horizontal direction seem to depend only on tongue biomechanics and virtual consonant target location, without any influence of the palate.

V. CONCLUSIONS

Simulations of VCV sequences (where C is a velar stop consonant) with a biomechanical model of the tongue have been presented. Both vowel and consonant gestures were controlled in terms of articulatory targets. Similar to observations on actual speakers, the VC and CV portions of the trajectories were somewhat curved and formed loops, even for symmetrical VCV sequences. The results seem to indicate that the presence and shape of the loops are strongly influenced by tongue biomechanics, including its muscular anatomy and contact with the palate. Contrary to suggestions by Löfqvist and Gracco (2002), biomechanics alone can be responsible for the trajectory curvature, and control of the entire trajectory based on a cost minimization principle does not seem to be necessary to explain these patterns. Of course, our results do not disprove Löfqvist and Gracco's (2002) hypothesis, since the control could act in combination with biomechanical factors. However, our simulations demonstrate that articulatory loops do not necessarily occur because entire articulatory trajectories are controlled in speech production. Our results support a more parsimonious theory of speech motor control, based on planning the target sequence and not the entire trajectory (Perrier *et al.*, 1996a).

For [uku] and [aka], the results of the simulations that depict forward-looping patterns are in good agreement with all the examples published in the literature. It was also shown for these two vowel contexts that in our simulations the direction and the orientation of the looping pattern is very resistant to changes in the position of the consonant target. Therefore, it seems that tongue biomechanics may explain the forward-oriented loop trajectories that were observed for these sequences on a number of different speakers and in different languages, while the upper portion of the loop is obviously influenced by interactions of the tongue with the palate.

Both for [iki] and [ika], in the first set of simulations, the model generated only backward movements. However, it was also shown for [iki] that a slight forward shift of the consonant target could induce a change in the loop direction and in its orientation. These results are consistent with the examples published in the literature. For example Houde (1967) observed for both sequences a small forward-looping pattern, but Mooshammer *et al.*'s (1995) findings were slightly different. First, the latter authors did not observe any looping pattern for their two speakers during the production of [ika]. Second, for [iki], they confirmed Houde's observation, but they also noted that the velocity at the onset of the closure was oriented rearward for one of their two speakers and forward for the other speaker. Therefore, it seems reasonable to assume that there is a certain amount of variability between speakers, and perhaps also between languages, in the orientation and shape of articulatory trajectories, when the vowel preceding the velar consonant is [i]. This characteristic seems to be properly accounted for by our model. However, it should also be noted that in our simulations it was never possible to generate the kind of forward loop that Houde (1967) observed for [ika], which is "*superimposed, during contact, on the main rearward movement of the tongue*" (p. 129).

In conclusion, the simulations reported in the current paper suggest that, whatever the vowel context, the articulatory patterns observed in VCV sequences, where C is a velar stop consonant, are largely determined by tongue biomechanics. However, especially in the case of [iki] and [ika], where the orientation of these patterns seems to be quite unstable, it is probably necessary to take into account the potential role of other factors, such as the precise locations of the consonant and vowel targets and aerodynamics. Preliminary studies of the fluid–walls interaction in the vocal tract lead us to infer that aerodynamics could have an influence when V1=[i] (Perrier *et al.*, 2000).

In general, our findings partially support Hoole *et al.*'s (1998) suggestion that both aerodynamics and biomechanics probably contribute to the generation of the loops: "*Taken together, these observations suggest that the elliptical movement patterns found in speech must be put down to at least two factors: Firstly, aerodynamic factors operating in the vicinity of a consonantal constriction; secondly, asymmetries in the muscle forces responsible for V-to-C and C-to-V movements.*" (Hoole *et al.*, 1998, p. 145). Our results may provide some answers to certain of Hoole's hypotheses. In particular, since in our simulations the commands patterns to all muscles are synchronized with each other, it may not be necessary to hypothesize temporal asymmetries in the muscle control to account for the generation of the observed loops. In addition, compared to the biomechanics, aerodynamics may have a limited influence, especially in back-vowel contexts.

ACKNOWLEDGMENTS

This work was supported by the CNRS, NSF, and NIH (NIDCD Grant DC DC01925). The authors thank Torstein Are, Vincent Coisy, Frédéric Jolly, and Jorge Vallejo, who contributed at different stages to the programming of the tongue model, Pierre Badin for the use of his x-ray data as a basis for the model and for valuable information about the 3D strain of the tongue and Christine Mooshammer and Susanne Fuchs for helpful comments at different stages of this work. Special thanks are also due to Anders Löfqvist and to the two reviewers, Phil Hoole and Khalil Iskarous, who provided extremely interesting and helpful comments.

[1]The equilibrium point hypothesis (EPH) and its associated λ model are at the center of a number of controversies in the motor control literature (for example, see Feldman and Levin, 1995; Gomi and Kawato, 1996; Gottlieb, 1998; Gribble *et al.*, 1998). Discussions about its possible use for the control of speech production have been presented in Laboissière *et al.* (1996), Perrier *et al.* 1996a, 1996b), and Sanguineti *et al.* (1998).

[2]Note that the amplitude of the sliding movement along the palate is identical for [ika] and [iki]. This is due simply to the fact that in both sequences the same targets were used for [i] and for [k]. Since a fairly long hold duration (100 ms) was specified for the consonant target command, no influence of the second vowel on the first part of the movement was to be expected. This is probably not very realistic, since, according to our hypothesis of speech production planning, the successive target commands should be selected specifically for each sequence, in order to optimize certain output constraints, such as economy of effort. This is a relatively minor detail that should not have a serious impact on the general conclusions drawn from our simulations.

Adams, S. G., Weismer, G., and Kent, R. D. (**1993**). ''Speaking rate and speech movement velocity profiles,'' J. Speech Hear. Res. **36**, 41–54.

Badin, P., Gabioud, B., Beautemps, D., Lallouache, T. M., Bailly, G., Maeda, S., Zerling, J. P., and Brock, G. (**1995**). ''Cineradiography of VCV sequences: Articulatory-acoustic data for a speech production model,'' in Proceedings of the 15th International Congress of Acoustics, Trondheim, Norway, Vol. IV, pp. 349–352.

Badin, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C., and Savariaux, C. (**2002**). ''Three-dimensional linear articulatory modeling of tongue, lips and face; based on MRI and video images,'' J. Phonetics **30**, 533–553.

Bothorel, A., Simon, P., Wioland, F., and Zerling, J.-P. (**1986**). ''Cinéradiographie des voyelles et des consonnes du français'' (Institut de Phonétique, Université Marc Bloch, Strasbourg, France).

Bunton, K., and Weismer, G. (**1994**). ''Evaluation of a reiterant force-impulse task in the tongue,'' J. Speech Hear. Res. **37**, 1020–1031.

Cooper, S. (**1953**). ''Muscle spindles in the intrinsic muscles of the human tongue,'' J. Physiol. (London) **122**, 193–202.

Dang, J., and Honda, K. (**1998**). ''Speech synthesis of VCV sequences using a physiological articulatory model,'' in Proceedings of the 5th International Conference on Spoken Language Processing, Sydney, Australia, Vol. 5, pp. 1767–1770.

Duck, F. A. (**1990**). *Physical Properties of Tissues: A Comprehensive Reference Book* (Academic, London).

Fant G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague), pp. 216–228.

Feldman, A. G. (**1966**). ''Functional tuning of the nervous system with control of movement or maintenance of a steady posture. II. Controllable parameters of the muscles,'' Biophysics (Engl. Transl.) **11**, 565–578.

Feldman, A. G. (**1986**). ''Once more on the equilibrium-point hypothesis (λ model) for motor control,'' J. Motor Behav. **18**(1), 17–54.

Feldman, A. G., and Orlovsky, G. N. (**1972**). ''The influence of different descending systems on the tonic stretch reflex in the cat,'' Exp. Neurol. **37**, 481–494.

Feldman, A. G., and Levin, M. F. (**1995**). ''Positional frames of reference in motor control: Origin and use,'' Behav. Brain Sci. **18**(4), 723–806.

Flanagan, J. R., Ostry, D. J., and Feldman, A. G. (**1990**). ''Control of human jaw and multi-joint arm movements,'' in *Cerebral Control of Speech and Limb Movements*, edited by G. E. Hammond (Elsevier Science, Amsterdam), pp. 29–58.

Fowler, C. A. (**1980**). ''Coarticulation and theories of extrinsic timing,'' J. Phonetics **8**, 113–133.

Fuchs, S., Perrier, P., and Mooshammer, C. (**2001**). ''The role of the palate in tongue kinematics: An experimental assessment in VC sequences from EPG and EMMA data,'' Proceedings of Eurospeech 2001, Aalborg, Denmark, pp. 1487–1490.

Fung, Y. C. (**1993**). *Biomechanics: Mechanical Properties of Living Tissues* (Springer, New York).

Gomi, H., and Kawato, M. (**1996**). ''Equilibrium-point control hypothesis examined by measured arm-stiffness during multi-joint movement,'' Science **272**, 117–120.

Gottlieb, G. L. (**1998**). ''Rejecting the equilibrium point hypothesis,'' Motor Control **2**, 10–12.

Gribble, P. L., Ostry, D. J., Sanguineti, V., and Laboissiere, R. (**1998**). ''Are complex control signals required for human arm movement?'' J. Neurophysiol. **79**(3), 1409–1424.

Hashimoto, K., and Suga, S. (**1986**). ''Estimation of the muscular tensions of the human tongue by using a three-dimensional model of the tongue,'' J. Acoust. Soc. Jpn. (E) **7**(1), 39–46.

Hoole, P., Munhall, K., and Mooshammer, C. (**1998**). ''Do air-stream mechanisms influence tongue movement paths?'' Phonetica **55**, 131–146.

Honda, K. (**1996**). ''The organization of tongue articulation for vowels,'' J. Phonetics **24**(1), 39–52.

Houde, R. A. (**1967**). ''A study of tongue body motion during selected speech sounds,'' Ph.D. Dissertation, University of Michigan (also published in 1968, in *SCRL Monograph No. 2*, Speech Comm. Res. Lab., Santa Barbara, California).

Kakita, Y., Fujimura, O., and Honda, K. (**1985**). ''Computation of mapping from muscular contraction patterns to formant patterns in vowel space,'' in *Phonetic Linguistics*, edited by V. A. Fromkin (Academic, Orlando, FL), pp. 133–144.

Keating, P. (**1993**). ''Fronted velars, palatalized velars, and palatals,'' Phonetica **50**, 73–101.

Kent, R., and Moll, K. (**1972**). ''Cinefluorographic analyses of selected lingual consonants,'' J. Speech Hear. Res. **15**, 453–473.

Kiritani, S., Miyawaki, K., and Fujimura, O. (**1976**). ''A computational model of the tongue,'' in *Annual Report of the Research Institute of Logopedics and Phoniatrics* (Tokyo University, Tokyo), Vol. 10, pp. 243–252.

Laboissière, R., Ostry, D. J., and Feldman, A. G. (**1996**). ''The control of multi-muscle systems: Human jaw and hyoid movements,'' Biol. Cybern. **74**(3), 373–384.

Löfqvist, A., and Gracco, V. L. (**1994**). ''Tongue body kinematics in velar stop production: Influences of consonant voicing and vowel context,'' Phonetica **51**, 52–67.

Löfqvist, A., and Gracco, V. L. (**1997**). ''Lip and jaw kinematics in bilabial stop consonant production,'' J. Speech Lang. Hear. Res. **40**, 877–893.

Löfqvist, A., and Gracco, V. L. (**2002**). ''Control of oral closure in lingual stop consonant production,'' J. Acoust. Soc. Am. **111**(6), 2811–2827.

Marhefka, D. W., and Orin, D. E. (**1996**). ''Simulations of contact using a non-linear damping model,'' in Proceedings of IEEE International Conference on Robotics and Automation, Minneapolis, MN, Vol. 2, pp. 1662–1668.

Matthies, M. L., Perrier, P., Perkell, J. S., and Zandipour, M. (**2001**). ''Variation in speech movement kinematics and temporal patterns of coarticulation with changes in clarity and rate,'' J. Speech Lang. Hear. Res. **44**(2), 340–353.

McClean, M. D., and Clay, J. L. (**1995**). ''Activation of lip motor units with variations in speech rate and phonetic structure,'' J. Speech Hear. Res. **38**, 772–782.

Min, Y., Titze, I., and Alipour, F. (**1994**). ''Stress–strain response of the human vocal ligament,'' NCVS Status Prog. Rep. **7**, 131–137.

Mooshammer, C., Hoole, P., and Kühnert, B. (**1995**). ''On loops,'' J. Phonetics **23**, 3–21.

Nelson, W. L. (**1983**). ''Physical principles for economies of skilled movements,'' Biol. Cybern. **46**, 135–147.

Netter, F. H. (**1989**). ''Atlas of human anatomy,'' CIBA-GEIGY Corporation editor.

Ohala, J. (**1983**). ''The origin of sound patterns in vocal-tract constraints,'' in *The Production of Speech*, edited by P. F. MacNeilage (Springer, New York), pp. 189–216.

Ohayon, J., Usson, Y., Jouk, P. S., and Cai, H. (**1999**). ''Fiber orientation in human fetal heart and ventricular mechanics: A small perturbation analysis,'' Comput. Methods Biomed. Eng. **2**, 83–105.

Öhman, S. E. G. (**1966**). ''Coarticulation in VCV utterances: Spectrographic measurements,'' J. Acoust. Soc. Am. **39**, 151–168.
Ostry, D. J., and Munhall, K. G. (**1985**). ''Control of rate and duration of speech movements,'' J. Acoust. Soc. Am. **77**, 640–648.
Payan, Y., and Perrier, P. (**1997**). ''Synthesis of V–V sequences with a 2D biomechanical tongue model controlled by the equilibrium point hypothesis,'' Speech Commun. **22**(2/3), 185–205.
Perkell, J. S. (**1969**). *Physiology of Speech Production: Results and Implication of a Quantitative Cineradiographic Study* (MIT Press, Cambridge, MA).
Perkell, J. S. (**1974**). ''A physiologically oriented model of tongue activity in speech production,'' unpublished Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
Perkell, J. S., and Matthies, M. L. (**1992**). ''Temporal measures of anticipatory labial coarticulation for the vowel [u]: Within- and cross-subject variability,'' J. Acoust. Soc. Am. **91**, 2911–2925.
Perkell, J. S., Svirsky, M. A., Matthies, M. L., and Manzella, J. (**1993**). ''On the use of electro-magnetic midsagittal articulometer (EMMA) systems,'' *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München, 31*, pp. 29–42 (University of Munich, Germany).
Perkell, J. S. (**1996**). ''Properties of the tongue help to define vowels categories: Hypotheses based on physiologically oriented modeling,'' J. Phonetics **24**, 3–22.
Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., and Zandipour, M. (**2000**). ''A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss,'' J. Phonetics **28**(3), 233–272.
Perkins, W. H., and Kent, R. D. (**1986**). *Functional Anatomy of Speech, Language and Hearing* (Allyn and Bacon, Needham Heights, MA).
Perrier, P., Loevenbruck, H., and Payan, Y. (**1996a**). ''Control of tongue movements in speech: The equilibrium point hypothesis perspective,'' J. Phonetics **24**, 53–75.
Perrier, P., Ostry, D. J., and Laboissière, R. (**1996b**). ''The equilibrium point hypothesis and its application to speech motor control,'' J. Speech Hear. Res. **39**, 365–378.
Perrier, P., Payan, Y., Perkell, J., Zandipour, M., Pelorson, X., Coisy, V., and Matthies, M. (**2000**). ''An attempt to simulate fluid–walls interactions during velar stops,'' in *Proceedings of the 5th Speech Production Seminar and CREST Workshop on Models of Speech Production* (Kloster Seeon, Bavaria), pp. 149–152.
Porter, R. (**1966**). ''Lingual mechanoreceptors activated by muscle twitch,'' J. Physiol. (London) **183**, 101–111.
Sanguineti, V., Laboissière, R., and Payan, Y. (**1997**). ''A control model of human tongue movements in speech,'' Biol. Cybern. **77**(11), 11–22.
Sanguineti, V., Laboissière, R., and Ostry, D. J. (**1998**). ''A dynamic biomechanical model for the neural control of speech production,'' J. Acoust. Soc. Am. **103**, 1615–1627.
Stone, M., Goldstein, M. H., and Zhang, Y. (**1997**). ''Principal component analysis of cross sections of tongue shapes in vowel production,'' Speech Commun. **22**, 173–184.
Straka, G. (**1965**). *Album Phonétique* (Les Presses de l'Université Laval, Québec, Canada).
Takemoto, H. (**2001**). ''Morphological analyses of the human tongue musculature for three-dimensional modeling,'' J. Speech Lang. Hear. Res. **44**, 95–107.
Trulsson, M., and Essick, G. K. (**1997**). ''Low-threshold mechanoreceptive afferents in the human lingual nerve,'' J. Neurophysiol. **77**, 737–748.
Walker, L. B., and Rajagopal, M. D. (**1959**). ''Neuromuscular spindles in the human tongue,'' Anat. Rec. **133**, 438.
Wilhelms-Tricarico, R. (**1995**). ''Physiological modeling of speech production: Methods for modeling soft-tissues articulators,'' J. Acoust. Soc. Am. **97**(5), 3085–3098.
Wood, S. A. J. (**1979**). ''A radiographic analysis of constriction locations for vowels,'' J. Phonetics **7**, 25–43.
Zienkiewicz, O. C., and Taylor, R. L. (**1989**). *The Finite Element Method. Basic Formulation and Linear Problems*. Maidenhead, UK (MacGraw-Hill, Maidenhead, UK).

The interlanguage speech intelligibility benefit[a)]

Tessa Bent[b)] and Ann R. Bradlow
Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208

(Received 20 December 2002; revised 6 June 2003; accepted 16 June 2003)

This study investigated how native language background influences the intelligibility of speech by non-native talkers for non-native listeners from either the same or a different native language background as the talker. Native talkers of Chinese ($n=2$), Korean ($n=2$), and English ($n=1$) were recorded reading simple English sentences. Native listeners of English ($n=21$), Chinese ($n=21$), Korean ($n=10$), and a mixed group from various native language backgrounds ($n=12$) then performed a sentence recognition task with the recordings from the five talkers. Results showed that for native English listeners, the native English talker was most intelligible. However, for non-native listeners, speech from a relatively high proficiency non-native talker from the same native language background was as intelligible as speech from a native talker, giving rise to the ''matched interlanguage speech intelligibility benefit.'' Furthermore, this interlanguage intelligibility benefit extended to the situation where the non-native talker and listeners came from different language backgrounds, giving rise to the ''mismatched interlanguage speech intelligibility benefit.'' These findings shed light on the nature of the talker–listener interaction during speech communication.
 [DOI: 10.1121/1.1603234]

PACS numbers: 43.71.Gv, 43.71.Hw, 43.71.Es [RD]

I. INTRODUCTION

It has long been recognized that adult second-language learners from a given native language background progress in a relatively consistent way from a monolingual to a bilingual system. Conversely, adult second-language learners from different native language backgrounds exhibit different deviations from the target language norms. For example, while native English talkers consistently produce and perceive the French high front rounded vowel [y] as [u] or [u]-like, native Portuguese talkers consistently produce and perceive French [y] as [i] or [i]-like (Rochet, 1995). This observation has formed the basis for a wealth of research and theorizing in phonetics and phonology on the perception and production of individual speech sound categories in second languages (e.g., see contributions to Strange, 1995). In this paper, we explore the phenomenon of native and target language interaction during the acquisition of second language sound structure at the level of overall sentence intelligibility. We investigate this issue by testing sentence intelligibility amongst adult non-native talkers and listeners who share a native language and amongst adult non-native talkers and listeners who do not share a native language. Since individuals from the same native language background who are in the process of acquiring a given target language all share an ''interlanguage,''[1] we predicted that target language intelligibility between non-native talkers and listeners from the same native language background would be enhanced relative to intelligibility between a native talker and a non-native listener. Indeed, second-language learners often report that the speech of a fellow non-native talker is easier to understand than the speech of a native talker. The present study aimed to investigate this claim under controlled laboratory conditions. Furthermore, we included talkers and listeners from several native language backgrounds in order to examine whether any ''interlanguage speech intelligibility benefit'' that we may observe (i.e., the benefit afforded by a shared interlanguage between a non-native talker and listener) is entirely dependent on the talker and listener sharing the native language or if non-native speech is generally more intelligible than native speech for all non-native listeners regardless of native language background.

Findings in support of these predictions would demonstrate that any measure of overall speech intelligibility must take into account both talker- and listener-related factors. Speech that is of relatively low intelligibility for one group of listeners (e.g., non-native accented English for native listeners of English) may be highly intelligible for another group of listeners (e.g., non-native listeners of English). Therefore, rather than considering variability in intelligibility as a function of talker characteristics, it is more accurate to measure intelligibility in terms of the talker–listener relationship. In this study, we approach this issue with respect to intelligibility between talkers and listeners from either the same or different native language backgrounds.

Previous research has demonstrated that native listeners generally find native talkers more intelligible than non-native talkers, particularly in noisy conditions (Munro, 1998; Munro and Derwing, 1995). However, the findings from studies investigating whether native or non-native talkers are more intelligible for non-native listeners have been less clear. For example, based only on casual observation rather than on controlled laboratory study, Nash (1969) claims that non-native speakers are more intelligible to other non-natives than to native speakers: ''A [non-native] speaker who cannot make himself understood when speaking English to a native English speaker will have no difficulty conversing in English

[a)]An earlier version of this study was presented at the 143rd meeting of the Acoustical Society of America, Pittsburgh, PA, June 2002.
[b)]Electronic mail: t-bent@northwestern.edu

with another [non-native] speaker'' (p. 4). Similarly, Weinreich (1953) states, ''When the other interlocutor is also bilingual, the requirements of intelligibility...are drastically reduced. Under such circumstances, there is hardly any limit to interference; forms can be transferred freely from one language to the other and often used in unadapted shape'' (p. 81). Smith and Rafiqzad (1979) provide some experimental evidence in support of this claim in a study in which foreign-accented talkers of English from multiple native language backgrounds and one native English talker recorded English passages of their own choosing. Then, listeners from various language backgrounds, including native languages that matched and mismatched the native languages of the talkers, completed a cloze procedure test in which the listeners filled in words that had been removed from transcripts of the recorded passages. Results showed that the non-native listeners found the native talker equally or less intelligible than the non-native talkers. Nevertheless, a talker–listener match with respect to native language did not necessarily produce the highest intelligibility scores. Unfortunately, the passages in this study differed significantly in difficulty and a strong positive correlation between passage difficulty and intelligibility was reported. This correlation suggests that the observed intelligibility differences may have been due to the difficulty of the passage rather than to the language backgrounds of the talkers and listeners.

More recently, van Wijngaarden (2001) and van Wijngaarden *et al.* (2002) provided strong evidence in support of the claim that under certain conditions non-native listeners find sentences produced by non-native talkers at least as intelligible as sentences produced by native talkers. Specifically, van Wijngaarden (2001) found that for native listeners, all four of the non-native talkers included in the study were less intelligible than each of the four native talkers included in the study. However, the two higher intelligibility non-native talkers (based on their ranking for the native listeners) were more intelligible for non-native listeners than each of the four native talkers. van Wijngaarden *et al.* (2002) measured intelligibility of Dutch, English, and German sentences produced by native and non-native talkers of these three languages for trilingual listeners (native language = Dutch, second language = English, third language = German). The listeners were all more proficient in English than in German as evidenced by the fact that they learned English earlier, used English more frequently, and gave higher self-reported scores on overall proficiency in English than in German. When listening to English (their higher proficiency foreign language) the listeners found the native English talkers more intelligible than the non-native talkers. However, when listening to German (their lower proficiency foreign language) they found the non-native talkers more intelligible than the native talkers. The findings of these studies demonstrate an intelligibility advantage between non-native talker–listener pairs when the listener is at a relatively early stage of target language acquisition and the talker is at a relatively advanced stage of target language acquisition. Similarly, Imai *et al.* (2003) found an advantage of shared language background for word recognition. They investigated the ability of native and non-native listeners with a first language of Spanish to recognize English words spoken by one native talker and one non-native talker also with Spanish as a first language. Native listeners performed better than the non-native listeners (i.e., recognized more words) for the words produced by the native talker, whereas the non-native listeners performed better than the native listeners for words produced by the non-native talker. An important finding from this study was that the non-native listeners outperformed the native listeners only for the words produced by the non-native talker that came from dense lexical neighborhoods (i.e., words that have many similar sounding ''neighbors'' with which they can easily be confused). This finding is consistent with the finding of Bradlow and Pisoni (1999) that word recognition by non-native listeners is worse for ''hard'' words (low-frequency words in high-density neighborhoods) than for ''easy'' words (high-frequency words in sparse neighborhoods). Since lexical neighborhood characteristics are defined in terms of segments shared across words, these findings suggest a connection between lower-level processing at the segmental level and higher-level processing at the word level.

Other work on non-native speech perception has shown that the success of acquiring the second language system depends on many factors, including (but not limited to) age of acquisition, duration of exposure, amount of continued first language usage, and quantity and quality of second language input (e.g., Flege, 2002). Although studies have shown that adults are capable of learning novel phonetic contrasts (Logan *et al.*, 1991; Lively *et al.*, 1993, 1994; Bradlow *et al.*, 1997), the impact of the native language has been shown to influence speech perception even for very early bilinguals. For example, Sebastian-Galles and Soto-Faraco (1999) showed that bilinguals who learned their second language as early as 4 years of age were less efficient at processing phonemic contrasts on-line than native listeners. The perception and production of non-native contrasts in late bilinguals and naive listeners is even more strongly influenced by the first language. However, an important generalization to emerge from studies of non-native phoneme perception is that not all non-native contrasts are equally difficult for second language learners. The relative difficulty with which non-native listeners perceive non-native sounds is related to the relationship between the status of sounds in the overall systems of phonemic contrasts in the first and second languages (e.g., Flege *et al.*, 1999; Best *et al.*, 2001). These difficulties in phoneme perception have been shown to affect lexical representation for early bilinguals (Pallier *et al.*, 2001) and to impair word recognition for early and late bilinguals (Meador *et al.*, 2000). Experiments with non-native listeners and native talkers have shown that non-native listeners can perform with native-like proficiency in the quiet, but even early bilinguals are impaired in noise compared to native listeners (e.g., Mayo *et al.*, 1997; McAllister, 1990). The cause for this decline is not fully understood, but a contributing factor may be the fact that non-native listeners are less able to take advantage of contextual cues for word identification (Mayo *et al.*, 1997), and so when in the presence of noise they have less signal-independent information to rely on as an aid to sentence recognition.

Two important issues regarding speech intelligibility between native and non-native talkers and listeners remain unresolved in controlled laboratory studies. First, all of the studies where both the talker and listener were non-native have involved a shared native language between talker and listener. The only exception is Smith and Rafiqzad (1979), where the uncontrolled variable of passage difficulty confounded the results. Therefore, we still do not know how the intelligibility of non-native talkers for non-native listeners varies depending on whether there is a match or mismatch in native language background between talker and listener. Data addressing this issue will allow us to assess the extent to which the possible non-native talker–listener intelligibility benefit is due to a particular shared interlanguage or to certain more general features of non-native language production and perception that are independent of the particular native and target languages in question. Second, the findings of van Wijngaarden (2001) and van Wijngaarden *et al.* (2002) suggested that talker and listener proficiency in the target language is important for determining the relative intelligibility of native and non-native talkers, but this factor deserves more attention. By including non-native talkers who are known to vary in target language production proficiency, the present study aimed to gain further insight into the factors that contribute to an interlanguage speech intelligibility benefit.

Using a database of sentence recordings by non-native talkers from various native language backgrounds, the present study investigated how the talker–listener match or mismatch with respect to language background and variation in second language proficiency affected non-native talker intelligibility. We made two specific predictions. First, we predicted, based on the findings of van Wijngaarden (2001) and van Wijngaarden *et al.* (2002), that a non-native talker with relatively high proficiency in English speech production would be at least as intelligible as a native talker for non-native listeners from the same native language background. This prediction supposes that non-native language learners from a given native language have a broad base of shared linguistic and phonetic knowledge that facilitates speech communication in the non-native language. We expected this interlanguage intelligibility benefit to be attenuated in the case of less proficient non-native talkers due to the fact that their speech production may stray so far from the target language norm that important lexical contrasts may not be adequately conveyed, resulting in word and sentence intelligibility that is low enough to produce very low overall intelligibility for both native and non-native listeners. Second, we predicted that a relatively high proficiency non-native talker will also be at least as intelligible as a native talker for non-native listeners from a different native language background. This prediction supposes that, regardless of native language background, certain features of non-native speech will make non-native talkers more intelligible to all non-native listeners. For example, non-native talkers may be less likely than native talkers to apply certain reduction phenomena that characterize native accented running speech in English such as alveolar flapping and failure to release final stops. While these features of non-native English speech may contribute to the overall impression of a foreign accent for native listeners, they may in fact be beneficial for speech perception by all non-native listeners of English regardless of native language background.

II. THE NORTHWESTERN UNIVERSITY FOREIGN ACCENTED ENGLISH SPEECH DATABASE

The Northwestern University Foreign Accented English Speech Database (NUFAESD) contains recordings of 64 sentences produced by 32 non-native talkers for a total of 2048 recorded sentences. Along with these recordings, the database includes demographic information about each talker and an overall intelligibility score for each talker as measured by a perception test with native English listeners.

The sentences in this database were taken from the sentence lists included in the revised Bamford–Kowal–Bench Standard Sentence Test (BKB-R) developed by the Cochlear Corporation for use with American children. The original Bamford–Kowal–Bench Standard Sentence Test was developed for use with British children (Bamford and Wilson, 1979; Bench and Bamford, 1979). These sentences were chosen for this database because they include words that are highly familiar to non-natives and are syntactically simple. Each list consists of 16 simple, declarative sentences with 3 or 4 keywords for a total of 50 keywords per list. From the 21 lists included in the BKB-R test, four lists (lists 7, 8, 9, and 10) were selected for the NUFAESD. These lists were selected based on their equivalent intelligibility scores for normal-hearing children as reported in Bamford and Wilson (1979). Ratings of age of acquisition, written and verbal frequency, imagability, concreteness, and familiarity for the 155 keywords in these four lists were gathered from the MRC Psycholinguistic Database: Machine Usable Dictionary Version 2.00 and are shown in Table I. Because some of the keywords are repeated within and across lists, there are not 200 unique keywords.

The talkers recorded for this database were recruited from the Northwestern University International Summer Institute and English as a Second Language (ESL) program over the course of the summer of 2000 and the 2000–2001 academic year. The International Summer Institute provides incoming international graduate students from across the university with intensive English language training as well as a general introduction to academic life in the USA during the month before they begin their graduate studies at Northwestern University. All of the participants in this program had already been admitted to a doctoral program and had therefore demonstrated a high level of proficiency with written English communication. However, these students all had limited experience with spoken English communication. The subjects recruited from the ESL program all came to the program due to their own or their department's recognition of their need to improve their oral and aural English skills. The talkers came from a variety of language backgrounds: Chinese[2] ($n=20$), Korean ($n=5$), Bengali ($n=1$), Hindi ($n=1$), Japanese ($n=1$), Romanian ($n=1$), Slovakian ($n=1$), Spanish ($n=1$), and Thai ($n=1$). The talkers ranged in age from 22 to 32 years with an average age of 25.5 years. They had spent on average 2.7 months in the United States

TABLE I. Frequency and psychological ratings for keywords.

	Verbal frequency[a]	Age of acquisition rating[b]	Written frequency[c]	Familiarity rating[d]	Imagability rating[d]	Concreteness rating[d]
Average	120	231	588	587	524	503
Standard deviation	395	58	2266	35	104	122
Number of items	134	37	134	106	104	97

[a]Verbal frequency: frequency of occurrence in verbal language from Brown (1984).
[b]Age of acquisition: age multiplied by 100 to produce a range from 100 to 700.
[c]Written frequency: the norms of Kucera–Francis (1967).
[d]Concreteness, familiarity, and imageability values range 100 to 700.

with a range from 0.25 months to 24 months at the time of recording. The average age at which the participants began to study English in school was 12.0 years and the average length of English learning was 9.8 years. All talkers had no known speech or hearing impairments. All talkers were paid for their participation.

The talkers were recorded in a sound-attenuated booth. The recording were made on an Ariel Proport A/D sound board with a Shure SM81 microphone. All subjects read the four sets of sentences in the same order. After the recording, the sound files were converted to the WAV format with a 16-kHz sampling rate and transferred to a PC-based computer. The digital speech files were then segmented into sentence length files and the root-mean-square amplitude was equated across all files.

For assessing the intelligibility of the talkers, the digitized recordings were embedded in white noise, yielding a speech-plus-noise file with a +5-dB signal-to-noise ratio. Each of the stimulus files consisted of a 400-millisecond silent leader, followed by 500 ms of noise, followed by the speech-plus-noise file, and ending with 500 ms of noise only.

In order to keep the number of native English listeners manageable while still evaluating the intelligibility of all the non-native talkers, eight test conditions with five listeners each were constructed. Each native English listener evaluated 64 sentences from the four BKB-R lists and all 32 non-native talkers described above. Each of the 32 non-native talkers supplied two different sentence recordings to each of eight test conditions (for a total of 64 sentences per condition). Therefore, each talker's intelligibility was evaluated on the basis of 16 sentences (8 conditions$\times$2 sentences per condition) from one full BKB-R list in a multiple talker presentation format (i.e., in each condition, listeners evaluated sentences by all 32 talkers). The native English listeners were presented with each sentence once and their task was to transcribe the sentence in standard English orthography. Intelligibility scores for each talker were based on the number of keywords correctly transcribed across all of the 16 sentences submitted to perceptual evaluation by the native English listeners. The raw percent-correct scores were converted to rationalized arcsine transform units (rau). This transformation places the scores on a linear and additive scale, thus facilitating meaningful statistical comparisons across the entire range of the scale (Studebaker, 1985). The rau scale ranges from -23 to 123. The middle of the scale corresponds closely to percent correct, but at the edges of the scale, below 15% and above 85%, the rau scale diverges from percent correct. For example, 80% corresponds to 79.9 rau but 95% corresponds to 102 rau. The non-native subjects' sentence production scores (i.e., the average percentage of keywords correctly transcribed by the native English listeners) ranged from 43 to 93 rau.

III. TEST OF FOREIGN-ACCENTED ENGLISH PERCEPTION

Using materials from the NUFAESD, we designed an experiment to investigate the perception of foreign-accented English by listeners from various language backgrounds. The overall design involved presenting English sentences produced by five talkers to four listener groups. The talkers were native talkers of Chinese ($n=2$), Korean ($n=2$), and English ($n=1$). The listeners were native talkers of Chinese ($n=21$), Korean ($n=10$), English ($n=21$), and a mixed group of non-native talkers of English from various native language backgrounds ($n=12$). None of the non-native listeners had served as talkers in the collection of the NUFAESD materials and none of the native English listeners had evaluated the sentences in the NUFAESD. Of particular interest for this study was the effect of a listener–talker match or mismatch in native language background, and the interaction of this listener–talker match or mismatch with talker proficiency in the target language (i.e., English).

A. Method

1. Talkers

The talkers selected for this experiment were from three language backgrounds: monolingual English ($n=1$), non-native (NN) talkers of English with a first language of Chinese ($n=2$), and NN talkers of English with a first language of Korean ($n=2$). The four non-native talkers' productions were taken from the NUFAESD. These talkers were selected based on their first language (either Korean or Chinese), gender (female), and production intelligibility scores obtained from the perceptual evaluation test conducted at the time of the database compilation. For the Chinese and Korean talkers, one from each language background was of higher proficiency and one of lower proficiency as defined by their production intelligibility scores. The Chinese high- and low-proficiency talkers had intelligibility scores (average percent of keywords correctly identified converted to the rau scale) of 80 and 43 rau, respectively. The Korean high- and low-

TABLE II. General information about non-native listener groups. Mean and standard deviations (in parentheses) are shown. The Northwestern University Graduate School minimum for TOEFL is 560 (paper) or 220 (computer).

	NN-Chinese ($n=21$)	NN-Korean ($n=10$)	NN-Mixed ($N=12$)	Significant differences
Age (years)	23.60 (1.60)	28.00 (4.72)	25.18 (2.35)	Korean>Chinese[a] Mixed>Chinese[b]
TOEFL[c]—computer	n/a	($n=1$) 253	($n=8$) 269.63 (22.06)	n/a
TOEFL—paper	($n=17$) 641.82 (19.63)	($n=7$) 602.43 (18.80)	($n=4$) 629.25 (37.25)	Chinese>Korean[a]
Time in US (months)	1.12 (1.92)	8.25 (9.48)	5.23 (10.34)	Korean>Chinese[a]
AOA[d] (years)	10.95 (2.22)	12.25 (1.39)	10.00 (4.22)	none
LOA[e] (years)	10.11 (2.26)	7.00 (1.41)	8.17 (4.37)	Chinese>Korean[a]

[a]$=p<0.01$.
[b]$=p<0.05$.
[c]TOEFL=Test of English as a Foreign Language.
[d]AOA=age of acquisition.
[e]LOA=length of acquisition (i.e., English study).

proficiency talkers had scores of 90 and 55 rau, respectively. An effort was made to match the intelligibility scores for the high-proficiency Chinese and Korean and the low-proficiency Chinese and Korean. However, no talkers in our sample with the selected demographic variables had identical production intelligibility scores. The talkers were not matched for any other features of their speech and we did not control for speaking rate in either the recording of the talkers or the selection of the talkers. The same procedure and equipment was used to record the monolingual English talker as was used in the compilation of the NUFAESD. No analogous intelligibility score is available for the native English talker since she was not included in any previous intelligibility testing.

2. Listeners

A total of 65 adults with normal speech and hearing participated in the experiment. The listeners came from four different language backgrounds: monolingual English ($n=21$), non-native speakers of English with a first language of Chinese (NN-Chinese, $n=21$), non-native speakers of English with a first language of Korean (NN-Korean, $n=10$), and non-native speakers of English with native language backgrounds other than Chinese or Korean (NN-mixed, $n=12$). The NN-mixed group included individuals from the following native language backgrounds: Bulgarian ($n=1$), Dutch ($n=1$), French/Douala ($n=1$), German ($n=1$), Greek ($n=2$), Hindi ($n=1$), Japanese ($n=2$), Serbian ($n=1$), Spanish ($n=1$), and Tamil ($n=1$). The non-native listeners were recruited from the Northwestern University International Summer Institute and ESL program over the course of the summer of 2001 and the 2001–2002 academic year. All listeners were paid for their participation in the study. Additional data for the three groups of non-native listeners are shown above in Table II.

The monolingual English listeners were all undergraduates at Northwestern University and ranged in age from 17 to 22 years with an average age of 19.1 years. The native English listeners were significantly younger than the non-native listeners [$t(59)=5.78$, $p<0.0001$]. They were recruited from the Linguistics Department subject pool and received course credit for their participation in the study. None of the listeners reported any speech or hearing problems.

3. Stimuli and procedure

Sentences from the four BKB-R lists (a total of 60 sentences) were divided into five lists of 12 sentences covering 37 keywords each (11 sentences with 3 keywords and 1 sentence with 4 keywords). See the Appendix for the sentence lists with the keywords underlined.

Subjects were seated in front of a computer monitor in a sound-attenuated booth. Stimulus presentation was controlled by special-purpose experiment running software (SUPERLAB PRO 2.01). The audio files were played out through the computer sound card (SoundBlaster Live) over headphones (Sennheiser HD 580). The subject's task was to listen to the sentence stimulus and to write down whatever she or he heard on specially prepared answer sheets. After each trial, the subject pressed a button on a response box (supplied as part of the SUPERLAB PRO 2.01 package) to elicit the next trial. Each trial was presented only once, but subjects could take as long as they needed to record their responses.

Each subject heard all five talkers and all five sentence lists in a blocked format. The monolingual English talker was always in the third position. The high-proficiency Chinese and Korean talkers were either first or second and the low-proficiency Chinese and Korean talkers were either fourth or fifth. The NN-Chinese and monolingual English listeners always heard the talkers in the following order: Chinese high proficiency, Korean high proficiency, native English, Chinese low proficiency, Korean low proficiency. For the NN-Korean listeners the order of the Chinese and Korean talkers was reversed. The rationale behind this ordering of the talkers was to ensure consistency across the NN-Chinese and NN-Korean groups with respect to the ordering of the listener–talker native language match and mismatch. Furthermore, the high-proficiency non-native talkers were ordered before the native talker so that superior performance

on the non-native talkers could not be attributed to a practice effect. Half of the NN-mixed listener group heard the lists in the order for the NN-Chinese group while the other half heard the lists in the order for the NN-Korean group. The particular sentence list read by each talker was counterbalanced across listeners.

In a separate session after the perception test, a word familiarity test was administered. For this test, each of the 144 unique keywords in the complete set of sentences used in this study was presented to the subjects for a familiarity rating on a scale of 1 to 7 where 1 = "I don't know this word," 4 = "I recognize this as an English word but I don't know its meaning," and 7 = "I know this word." A set of 75 filler items was also presented as part of this test. These filler items were selected from lists of words that were given low, medium, and high familiarity ratings by native listeners in Lewellen *et al.* (1993) and that were used in previous tests with both native and non-native listeners (Bradlow and Pisoni, 1999). The inclusion of these words ensured that the full range of the familiarity scale would be represented by the items in this test. An additional 128 words were also included in this test for the purpose of a separate study not reported here. On each trial, the target word was presented in standard American English orthography on the computer screen (using SUPERLAB PRO 2.01 software), and the subject entered his or her familiarity rating by pressing the appropriate button on the keyboard. The item remained on the screen until a response was recorded, which then triggered the start of the next trial. The order of presentation of the items was randomized.

4. Data analysis

Sentence-in-noise perception scores were determined by a strict keyword-correct count. For each set of sentences heard by each listener, the talker could receive a score from 0 to 37 keywords correct. This score was obtained by counting the number of keywords transcribed perfectly. Words with added or deleted morphemes were counted as incorrect. However, obvious spelling errors were not counted as incorrect. Raw intelligibility scores were converted to percent correct and then to rationalized arcsine units (rau).

B. Results

1. Word familiarity

Data from the word-familiarity rating task showed that the vast majority of words was highly familiar to the vast majority of the non-native listeners. Familiarity data from two of the 44 listeners were missing: one did not return for the second data collection session when the word-familiarity test was administered and the other's data had to be discarded due to a computer error. Of the remaining 42 listeners, all gave high ratings (5 or greater) to at least 94% of the words. Thirty-two listeners gave scores of less than 5 to no more than two words; six listeners gave scores of less than 5 to three to five words; four listeners gave scores of less than 5 to six to nine words, and no listeners gave scores of less than 5 to more than nine words. Furthermore, there was no correlation between the listeners' average familiarity rating score and their average score on the sentence-in-noise perception test ($\text{rho} = 0.120$, $p = 0.47$). A vast majority of the words, 79%, was given a rating of 7 by all subjects. Additionally, only five words had average scores lower than 6: broom (5.8), buckets (5.8), janitor (5.6), jug (5.1), and saucepan (5.9). Of the 144 target items, only 19 words were given any scores lower than 5. Nine words were given a score under 5 by one listener; three words were given scores under 5 by two listeners; and seven words were given scores lower than 5 by more than two listeners (from 5 to 18 listeners). Last, the non-natives' scores on the low-, mid-, and high-familiarity filler items were similar to scores given by native listeners in Bradlow and Pisoni (1999). The low-, mid-, and high-familiarity filler items were given scores of 1.83, 3.88, and 6.93, respectively, by the native listeners in Bradlow and Pisoni's study and were given scores of 2.85, 3.79, and 6.53, respectively, by the non-native listeners in the current study. Therefore, we performed all analyses of the sentence-in-noise perception test with the assumption that the non-native listeners were all sufficiently familiar with the keywords to ensure that this test provided a valid measure of their ability to perceive sentences in noise independently of word familiarity.

2. Foreign-accented English perception

Table III summarizes the five talkers' intelligibility scores for each of the four listener groups.

A repeated measures ANOVA with listener (native English, NN-Chinese, NN-Korean, NN-mixed) as the between-subjects factor and talker (high-proficiency Chinese, low-proficiency Chinese, high-proficiency Korean, low-proficiency Korean, native English) as the within-subjects factor showed highly significant main effects of listener [$F(3,240) = 39.34$, $p < 0.0001$] and talker [$F(4,240) = 194.15$, $p < 0.0001$]. The interaction of talker and listener was highly significant [$F(12,240) = 11.37$, $p < 0.0001$]. For each listener group the low-proficiency Chinese and Korean talkers were less intelligible than the native talker and the high-proficiency Chinese and Korean talkers. However, the rankings for the two high-proficiency non-natives and the native talker depended on the language background of the listener group.

Post hoc pairwise comparisons (Bonferroni/Dunn tests) of talker intelligibility within each listener group were conducted. Due to the large number of paired comparisons (ten for each listener group), the p value must be less than 0.005 to be significant. Pairwise comparisons within a talker across listener groups were not conducted as those comparisons were not of primary interest for this study. For the NN-Chinese listeners, the high-proficiency Chinese talker was not significantly different from the high-proficiency Korean talker or the native English talker. However, the high-proficiency Korean talker was significantly more intelligible than the native English talker ($p = 0.001$). All three of these talkers were significantly more intelligible for these listeners than the low-proficiency Chinese and Korean talkers ($p < 0.001$). For the NN-Korean listeners, there were no significant differences between the high-proficiency Chinese, high-proficiency Korean, and native English talkers. Addi-

TABLE III. Percent keywords correctly transcribed for talkers and listener groups in rau. Standard deviations are shown in parentheses. Scores in bold are significantly higher than the other scores in the row and are not significantly different from one another except for the underlined scores. Specifically, for the NN-Chinese listeners, the high-proficiency Korean talker is significantly more intelligible than the native English, and for the NN-Korean listeners the high-proficiency Korean talker is significantly more intelligible than the low proficiency Korean talker.

	Talker					
Listener group	Chinese high proficiency	Korean high proficiency	Native English	Chinese low proficiency	Korean low proficiency	All talkers
NN-Chinese ($n=21$)	**64** (**10.8**)	**66** (**11.7**)	**56** (**10.4**)	30 (12.1)	41 (9.8)	51 (17.4)
NN-Korean ($n=10$)	**60** (**15.5**)	**74** (**15.8**)	**60** (**11.7**)	22 (11.4)	**53** (**12.0**)	54 (21.7)
NN-Mixed ($n=12$)	**62** (**11.0**)	**70** (**7.7**)	**67** (**15.8**)	19 (20.9)	41 (14.8)	52 (24.1)
Native English ($n=21$)	77 (12.2)	91 (8.4)	**109** (**14.7**)	38 (13.8)	60 (12.1)	75 (27.7)
All listeners	67 (13.9)	76 (15.2)	76 (26.8)	29 (15.9)	49 (14.5)	

tionally, the high-proficiency Chinese and native English talkers were not significantly different from the low-proficiency Korean talker. However, the high-proficiency Korean talker was significantly more intelligible than the low-proficiency Korean talker ($p<0.001$). For the NN-mixed listener group, the high-proficiency Chinese, high-proficiency Korean, and the native English talkers were not significantly different from one another, and were all significantly more intelligible than the low-proficiency Chinese and low-proficiency Korean talkers ($p<0.0001$). For the native English listeners, the native English talker was significantly more intelligible than all the other talkers ($p<0.001$).

In summary, the native listeners found the native talker more intelligible than any of the other talkers, and all listeners generally found the lower-proficiency non-native talkers less intelligible than either their high-proficiency counterparts or the native talker. Of particular interest for this study was the finding that the non-native listeners found the high-proficiency non-native talkers with whom they shared native language as intelligible as the native English talker. This finding also extended, in one case, to a low-proficiency non-native talker in that the low-proficiency Korean talker was as intelligible as the native English talker for the NN-Korean listener group. This finding demonstrates a *matched interlanguage speech intelligibility benefit*, such that a native language match between a non-native talker and a non-native listener facilitates speech intelligibility. Furthermore, when the non-native listeners and high-proficiency non-native talkers did not share a native language, the non-native listeners found the non-native talkers equally as or more intelligible than the native talker. This general finding suggests that the interlanguage benefit can extend to the situation of a talker–listener native language mismatch, demonstrating a *mismatched interlanguage speech intelligibility benefit*.

To investigate the possible contribution of speaking rate to the intelligibility results, we measured sentence duration for each of the five speakers. All pairwise comparisons amongst the five talkers were significant ($p<0.005$) except for the high-proficiency Korean and the low-proficiency Chinese talker, whose average sentence durations did not differ significantly. The average sentence durations for the five talkers are shown in Table IV.

For each of the non-native listener groups, the average sentence durations for the five talkers did not significantly correlate with their intelligibility scores (Spearman rank correlation, NN-Chinese listeners: rho$=0.300$, $p=0.55$; NN-Korean listeners: rho$=0.103$, $p=0.84$; NN-Mixed listeners: rho$=-0.100$, $p=0.84$; Native English listeners: rho $=-0.300$, $p=0.55$). For example, the low-proficiency Korean talker had longer sentence durations than the high-proficiency Korean, yet almost all the non-native listeners (except for three NN-Korean listeners) found the high-proficiency Korean more intelligible than the low-proficiency Korean. This analysis suggests that the observed interlanguage intelligibility benefit is not simply due to variability in speaking rate across the talkers.

IV. SUMMARY AND DISCUSSION

Four groups of listeners (monolingual English, NN-Chinese, NN-Korean, and NN-mixed) transcribed sentences produced by a native English talker, two Chinese non-native talkers of English, and two Korean non-native talkers of English. The major finding of this study was that the relative intelligibility of each talker depended on the language background of the listener such that

TABLE IV. Average sentence durations for the five talkers in order of increasing duration. Means in milliseconds and standard deviations (in parentheses) are shown.

	Average sentence duration
Native English	1223 ms (161)
Low-proficiency Chinese	1512 ms (244)
High-proficiency Korean	1587 ms (210)
Low-proficiency Korean	1680 ms (280)
High-proficiency Chinese	1717 ms (196)

(1) For native listeners, intelligibility of the native talker was greater than the intelligibility of any of the non-native talkers.
(2) For non-native listeners, intelligibility of a high-proficiency non-native talker (and in one case a low-proficiency talker) from the same native language background was equal to the intelligibility of the native talker. This is the ''matched interlanguage speech intelligibility benefit.''
(3) For non-native listeners, intelligibility of a high-proficiency non-native talker from a different native language background was greater than or equal to the intelligibility of the native talker. This is the ''mismatched interlanguage speech intelligibility benefit.''

The matched interlanguage speech intelligibility benefit can be explained by the fact that non-native speech production and perception are both systematically linked to native language sound structure (for a wealth of supporting research and theory see Strange, 1995; Best, 1994, 1995; Flege, 1992, 1995; Kuhl and Iverson, 1995 and many others). Thus, the speech of a non-native talker is more intelligible to non-native listeners with whom they share a native language than for native listeners due to the fact that the overall shared phonetic and phonological knowledge between the non-native talker and non-native listeners from the same language background is likely to be more extensive than a native/non-native pair. For the non-natives who share a native language, their linguistic knowledge covers aspects of both the native and target languages, whereas for the non-native/native pair the shared knowledge base includes only their knowledge of the target language insofar as it is developed in the non-native talker. This shared knowledge base includes the system of consonant and vowel categories, phonotactics, stress patterns, and intonation as well as other features of the sound system. Thus, a non-native listener is well equipped to interpret certain acoustic–phonetic features of the speech of a native-language-matched non-native talker as the talker intended them to be interpreted, even though they may deviate markedly from the target language norm, whereas native listeners are better equipped to interpret the speech of a native talker. For example, even foreign-accented talkers who have gained control over producing the full inventory of vowel contrasts of the target language may produce the vowels of the target language in the region of the vowel space (i.e., with a base of articulation) that is typical of the native language rather than of the target language. While this may result in misinterpretation of a particular vowel for native listeners or non-native listeners from a different native language background, non-native listeners from the same native language background as the talker will be more likely to access the correct vowel category, thereby contributing to the matched interlanguage speech intelligibility benefit that we observed in this study.

A possible explanation for the mismatched interlanguage speech intelligibility benefit is that it results from the talker's and listener's shared knowledge of the structure of the target language in conjunction with the influence of general strategies that listeners and talkers apply when learning to produce and perceive a foreign language. For example, in native-accented American English word-final stop consonants are frequently unreleased. Native listeners know to listen for cues to the presence and identity of word-final consonants in other parts of the signal and to interpret the lack of a salient word-final stop release as reflecting structural and/or contextual influences at the phrase and discourse levels, but non-native listeners may miss these cues, thereby compromising their overall comprehension of native-accented English. However, in their own English speech, non-native talkers who have not yet mastered all the details of American English allophony may produce particularly salient word-final stop consonant releases, thereby facilitating the intelligibility of their speech for other non-native listeners from a wide range of native language backgrounds (though not necessarily for native listeners). For example, Smith, Bradlow, and Bent (2003) demonstrated that non-native listeners from various native language backgrounds are better at identifying words in minimal pairs that differ only in the voicing of the final stop (e.g., cap vs cab, pick vs pig) produced by non-native talkers than produced by native talkers. In this case, even though the native listeners generally performed better than the non-native listeners, the performance of the non-native listeners surpassed that of the native listeners for words produced by one non-native talker, indicating that the non-native listeners must be listening for certain cues that native listeners are not attending to.

Alternatively, the mismatched interlanguage speech intelligibility benefit that we observed in this study may stem from similarities in the sound structure of the two languages that we investigated, in which case it is in fact just another manifestation of the matched interlanguage benefit rather than a separate phenomenon. For example, both Chinese and Korean have a much more constrained syllable structure than English, including a constraint against final consonant clusters. The transfer from a language which does not allow final consonant clusters such as Chinese or Korean to a language that does allow coda clusters such as English may result in similar features of Chinese-accented and Korean-accented English which serve to facilitate perception of English speech between native Chinese and native Korean listeners. However, evidence against this account based on structural similarities between Chinese and Korean comes from the results of the test with the NN-mixed listener group which included very small number of listeners (only 1 or 2) from each of a wide range of language backgrounds. Seven of these 12 listeners found the high-proficiency Chinese talker equally or more intelligible than the English talker, and five of the 12 listeners from this highly heterogeneous listener group found the high-proficiency Korean talker more intelligible than the native English talker. These listeners (i.e., those from the NN-mixed group who showed a mismatched interlanguage speech intelligibility benefit) came from a diverse group of native language backgrounds including Bulgarian, Dutch, French/Douala (bilingual), Greek, Hindi, Japanese, and Spanish. This finding suggests that the observed mismatched interlanguage speech intelligibility benefit is unlikely to be the result of structural similarities between the native languages of the talkers and listeners, but

rather due to certain tendencies in foreign-accented English regardless of native language background. Nevertheless, data from studies with additional carefully selected languages are needed in order to rule out this alternative account.

From the current study, we cannot determine at which stage of spoken language processing the interlanguage benefit arises since perceiving words in sentences requires processing on many different levels and our task was an off-line measure of comprehension that represents the accumulation of processing at multiple levels. Because we controlled for the linguistic content of the sentences, we can assume that the advantage that many of the non-native listeners received when listening to non-native speech came from the differences in the acoustic signal and not from differences in lexical choices or syntactic structures. Therefore, it is very likely that the source of the observed interlanguage effect was at a relatively early, phonetic stage of processing. Nevertheless, it is possible that the interlanguage benefit also operates at higher levels of sentence comprehension where other aspects of linguistic structure and of extralinguistic factors come into play. Additional tests that specifically tap into various levels of spoken language processing will be required to determine whether the early processes are most important for the interlanguage benefit or if the localization of the phenomenon occurs later in the processing of spoken language as well.

Large individual differences were found in the magnitude of the interlanguage benefit. The difference in intelligibility between the high-proficiency non-native talker and the native talker ranged from −23 to 52 rau for the matched interlanguage benefit and from −24 to 39 rau for the mismatched interlanguage benefit. The basis of these large individual differences remains unknown. The subjects in this study were rather homogeneous in terms of demographic variables (e.g., age of English study onset, length of residence, etc.) and in terms of English proficiency [e.g., TOEFL scores ranged only from 573–677 (paper) and 220–290 (computer)]. Future research will be needed to determine the source of these differences. For example, independent tests of language proficiency, particularly with respect to target language speech perception, will help to determine how listener proficiency in the non-native language influences the preference for non-native over native speech.

The findings from the present study are consistent with the findings of van Wijngaarden (2001) and van Wijngaarden *et al.* (2002), which demonstrated a non-native talker speech intelligibility advantage for non-native listeners at a relatively early stage of target language acquisition. Two important methodological differences between the present study and the studies by van Wijngaarden and colleagues are the target language (English vs Dutch) and the intelligibility measurement technique: van Wijngaarden and colleagues measured the signal-to-noise ratio required for a 50%-correct response (the speech reception threshold, or SRT), whereas the present study measured intelligibility in terms of a keyword recognition accuracy score for sentences presented at a fixed signal-to-noise ratio. Taken together, the present study and those of van Wijngaarden and colleagues provide converging evidence for an interlanguage speech intelligibility benefit and demonstrate that any measure of speech intelligibility must take into account both talker- and listener-related factors.

We conclude by noting an implication of the present findings for language variation and change. The demonstration of the interlanguage speech intelligibility benefit suggests a mechanism that may underlie the establishment of new pronunciation norms across a community of non-native talkers. Due to various social, political, and historical circumstances, non-native talkers who share a native language may sometimes communicate in a shared foreign language. This situation can arise in settings where the shared foreign language dominates in the broader context. For example, in many university research laboratories where the director and experimenters share a native language, the typical language of the laboratory may be English due to the fact that the broader scientific community is English dominated. This situation can occur on an even larger scale such as in a country like India, where English is widely spoken as a second language and, in certain settings, even talkers from the same native language background will communicate in English. Under such circumstances, a characteristic variety of the target language (e.g., Indian English) may develop as a result of the interlanguage speech intelligibility benefit which will cause certain interlanguage features to become firmly entrenched in the speech of the community. This intelligibility-based factor will likely operate in conjunction with other sociolinguistic factors to reinforce the establishment of a new and lasting variety of the target language.

ACKNOWLEDGMENTS

We are grateful to Martin Armstrong, Lyla Miller, Mengting Shieh, and Saundra Wright for data collection assistance. This research was supported by NIH-NIDCD Grant No. DC 03762.

APPENDIX: FOREIGN-ACCENTED ENGLISH PERCEPTION TEST MATERIALS

Based on BKB-R materials (Bamford and Wilson, 1979; Bench and Bamford, 1979).

List 1

(1) The children dropped the bag.
(2) The dog came back.
(3) The floor looked clean.
(4) She found her purse.
(5) The fruit is on the ground.
(6) Mother got a saucepan.
(7) They washed in cold water.
(8) The young people are dancing.
(9) The bus left early.
(10) The ball is bouncing very high.
(11) Father forgot the bread.
(12) The girl has a picture book.

List 2

(1) The boy forgot his book.
(2) A friend came for lunch.
(3) The match boxes are empty.

(4) He climbed his ladder.
(5) The family bought a house.
(6) The jug is on the shelf.
(7) The ball broke the window.
(8) They are shopping for cheese.
(9) The pond water is dirty.
(10) They heard a funny noise.
(11) The police are clearing the road.
(12) The bus stopped suddenly.

List 3

(1) The book tells a story.
(2) The young boy left home.
(3) They are climbing the tree.
(4) She stood near her window.
(5) The table has three legs.
(6) A letter fell on the floor.
(7) The five men are working.
(8) The shoes were very dirty.
(9) They went on a vacation.
(10) The baby broke his cup.
(11) The lady packed her bag.
(12) The dinner plate is hot.

List 4

(1) A dish towel is by the sink.
(2) She looked in her mirror.
(3) The good boy is helping.
(4) They followed the path.
(5) The kitchen clock was wrong.
(6) Someone is crossing the road.
(7) The mailman brought a letter.
(8) They are riding their bicycles.
(9) He broke his leg.
(10) The milk was by the front door.
(11) The shirts are hanging in the closet.
(12) The chicken laid some eggs.

List 5

(1) The orange was very sweet.
(2) He is holding his nose.
(3) The new road is on the map.
(4) She writes to her brother.
(5) The football player lost a shoe.
(6) The three girls are listening.
(7) The coat is on a chair.
(8) The train is moving fast.
(9) The child drank some milk.
(10) The janitor used a broom.
(11) The ground was very hard.
(12) The buckets hold water.

[1]Selinker (1972) defined an interlanguage as ''a separate linguistic system based on the observable output which results from a learner's attempted production of a target language norm'' (p. 214).
[2]All Chinese subjects were native speakers of Mandarin Chinese in the sense that all schooling including University had been in Mandarin. However, many of the Chinese subjects spoke another dialect of Chinese at home.

Bamford, J., and Wilson, I. (**1979**). ''Methodological considerations and practical aspects of the BKB sentence lists,'' in *Speech-hearing Tests and the Spoken Language of Hearing-impaired Children*, edited by J. Bench and J. Bamford (Academic, London), pp. 148–187.

Bench, J., and Bamford, J., Editors (**1979**). *Speech-hearing Tests and the Spoken Language of Hearing-impaired Children* (Academic, London).

Best, C. T. (**1994**). ''The emergence of native-language phonological influences in infants: A perceptual assimilation model,'' in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by J. Goodman and H. C. Nusbaum (MIT Press, Cambridge, MA), pp. 167–224.

Best, C. T. (**1995**). ''A direct realist view of cross-language speech perception,'' in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 171–204.

Best, C., McRoberts, G., and Goodell, E. (**2001**). ''Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system,'' J. Acoust. Soc. Am. **109**, 775–794.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., and Tohkura, Y. (**1997**). ''Training Japanese listeners to identify English /r/ and /l/. IV. Some effects of perceptual learning on speech production,'' J. Acoust. Soc. Am. **101**, 2299–2310.

Bradlow, A. R., and Pisoni, D. B. (**1999**). ''Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors,'' J. Acoust. Soc. Am. **106**, 2074–2085.

Brown, G. D. A. (**1984**). ''A frequency count of 190,000 words in the London-Lund corpus of English conversation,'' Behav. Res. Methods Instrum. **16**(6), 502–532.

Flege, J. E. (**1992**). ''Speech learning in a second language,'' in *Phonological Development: Models, Research, and Applications*, edited by C. Ferguson, L. Menn, and C. Steol-Gammon (York, Timonium, MD), pp. 565–604.

Flege, J. E. (**1995**). ''Second language speech learning: Theory, findings, and problems,'' in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 233–277.

Flege, J. E. (**2002**). ''Interactions between the native and second-language phonetics systems,'' in *An Integrated View of Language Development: Papers in Honor of Henning Wode*, edited by P. Burmeister, T. Piske, and A. Rohde (Trier, Wissenschaftlicker Trier).

Flege, J. E., MacKay, I. R. A., and Meador, D. (**1999**). ''Native Italian speakers' perception and production of English vowels.'' J. Acoust. Soc. Am. **106**, 2973–2987.

Imai, S., Flege, J. E., and Walley, A. (**2003**). ''Spoken word recognition of accented and unaccented speech: Lexical factors affecting native and non-native listeners,'' in Proceedings of the International Congress on Phonetic Science, Barcelona, Spain.

Kucera, H., and Francis, W. N. (**1967**). *Computational Analysis of Present-Day American English* (Brown University Press, Providence).

Kuhl, P. K., and Iverson, P. (**1995**). ''Linguistic experience and the perceptual magnet effect,'' in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 121–154.

Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., and Greene, B. G. (**1993**). ''Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision and semantic categorization.'' J. Exp. Psychol. Gen. **122**, 316–330.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). ''Training Japanese listeners to identify English /r/ and /l/. II. The role of phonetic environment and talker variability in learning new perceptual categories,'' J. Acoust. Soc. Am. **94**, 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (**1994**). ''Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories,'' J. Acoust. Soc. Am. **96**, 2076–2087.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). ''Training Japanese listeners to identify English /r/ and /l/: A first report,'' J. Acoust. Soc. Am. **89**, 874–886.

Mayo, L. H., Florentine, M., and Buus, S. (**1997**). ''Age of second-language acquisition and perception of speech in noise,'' J. Speech Lang. Hear. Res. **40**, 686–693.

McAllister, R. (**1990**). ''Perceptual foreign accent: L2 user's comprehension ability,'' in Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second-language Speech, pp. 216–228.

Meador, D., Flege, J. E., and MacKay, I. R. A. (**2000**). ''Factors affecting the

recognition of words in a second language,'' Bilingualism: Lang. Cogni. **3**, 55–67.
Munro, M. J., and Derwing, T. M. (**1995**). ''Foreign accent, comprehensibility, and intelligibility in the speech of second language learners,'' Lang. Learn. **45**, 73–97.
Munro, M. J. (**1998**). ''The effects of noise on the intelligibility of foreign-accented speech,'' Stud. Second Lang. Acquisit. **20**, 139–153.
Nash, R. (**1969**). ''Intonational interference in the speech of Puerto Rican bilinguals,'' J. English **4**, 1–42.
Pallier, C., Colome, A., and Sebastian-Galles, N. (**2001**). ''The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries,'' Psychol. Sci. **12**, 445–449.
Rochet, B. (**1995**). ''Perception and production of second-language speech sounds by adults,'' in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 379–410.
Sebastian-Galles, N., and Soto-Faraco, S. (**1999**). ''Online processing of native and non-native phonemic contrasts in early bilinguals,'' Cognition **72**, 111–123.
Selinker, L. (**1972**). ''Interlanguage.'' IRAL **10**, 209–231.
Smith, B., Bradlow, A. R., and Bent, T. (**2003**). ''Production and perception of temporal contrasts in foreign-accented English,'' in Proceedings of the International Congress on Phonetic Science, Barcelona, Spain.
Smith, L. E., and Rafiqzad, K. (**1979**). ''English for cross-cultural communication: The question of intelligibility,'' TESOL Quarterly **13**, 371–380.
Strange, W., Editor (**1995**). *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (York, Baltimore).
Studebaker, G. A. (**1985**). ''A rationalized arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.
van Wijngaarden, S. J. (**2001**). ''Intelligibility of native and non-native Dutch speech.'' Speech Commun. **35**, 103–113.
van Wijngaarden, S. J., Steeneken, H. J. M., and Houtgast, T. (**2002**). ''Quantifying the intelligibility of speech in noise for non-native listeners.'' J. Acoust. Soc. Am. **111**, 1906–1916.
Weinreich, U. (**1953**). *Languages in Contact: Findings and Problems* (Mouton, The Hague).

Stimulus presentation order and the perception of lexical tones in Cantonese[a)]

Alexander L. Francis[b)]
Department of Audiology and Speech Sciences, Purdue University, West Lafayette, Indiana 47907

Valter Ciocca[c)]
Department of Speech and Hearing Sciences, University of Hong Kong

(Received 9 August 2002; revised 18 May 2003; accepted 30 June 2003)

Listeners' auditory discrimination of vowel sounds depends in part on the order in which stimuli are presented. Such presentation order effects have been argued to be language independent, and to result from psychophysical (not speech- or language-specific) factors such as the decay of memory traces over time or increased weighting of later-occurring stimuli. In the present study, native Cantonese speakers' discrimination of a linguistic tone continuum is shown to exhibit order of presentation effects similar to those shown for vowels in previous studies. When presented with two successive syllables differing in fundamental frequency by approximately 4 Hz, listeners were significantly more sensitive to this difference when the first syllable was higher in frequency than the second. However, American English-speaking listeners with no experience listening to Cantonese showed no such contrast effect when tested in the same manner using the same stimuli. Neither English nor Cantonese listeners showed any order of presentation effects in the discrimination of a nonspeech continuum in which tokens had the same fundamental frequencies as the Cantonese speech tokens but had a qualitatively non-speech-like timbre. These results suggest that tone presentation order effects, unlike vowel effects, may be language specific, possibly resulting from the need to compensate for utterance-related pitch declination when evaluating fundamental frequency for tone identification.
[DOI: 10.1121/1.1603231]

PACS numbers: 43.71.Hw [PFA]

I. INTRODUCTION

The discriminability of speech sounds has been shown to depend, in some cases, on the order in which stimuli are presented. For example, Repp, Healey, and Crowder (1979) described the results of experiments in which listeners were asked to judge the similarity of pairs of vowels selected from a continuum ranging from [i] to [I]. For a given pair of stimuli, when the initial vowel was more [i]-like, listeners tended to respond ''same'' significantly more often than when the order was reversed. Similar results were reported by Cowan and Morse (1986), and Macmillan, Braida, and Goldberg (1987). All of these authors suggested that such order effects might result from the decay of the memory trace of the initial token of the pair. According to this explanation, the memory trace of a vowel decays ''toward'' a vowel that is more centrally located in an abstract representation of the talker's vowel space (e.g., [ə]). This model could be termed a ''neutralization hypothesis'' because the memory trace becomes more like a neutral vowel (one that is not particularly high or low, front or back in the vowel space) in a manner similar to the reduction or neutralization of vowel quality in the production of unstressed English vowels (cf. Ladefoged, 2001, pp. 78–79). Thus, along an [i]–[ɪ] continuum, more [i]-like vowels, decaying toward [ə], become more [ɪ]-like, increasing the perceived similarity of the pair. When the initial token of the pair is more [ɪ]-like, its memory trace also decays toward [ə], but in this case that is ''away'' from [i] in the vowel space, leading to a decrease in the perceived similarity of the pair. There are two hypotheses regarding the cause of this observed tendency. One exploits the specific geometry of vowel spaces and one does not draw on linguistic factors at all.

Cowan and Morse (1986) argued that the direction and extent of decay of [i]-like stimuli is determined by the boundaries of the listeners' vowel space. However, their theory does not specify how these boundaries are defined. It is possible that they are determined entirely by the listener's experience, such that those previously encountered tokens with extreme first formant ($F1$) and second formant ($F2$) values determine the boundaries at any given time. Alternatively, these boundaries could be defined in an experience-independent (innate) manner based on the limits of the interaction of articulatory and auditory systems. As discussed by Lieberman and Blumstein (1988, pp. 171–183), the extreme articulatory configurations of [i], [ɑ], and [u] both impart significant acoustic stability to these vowels (cf. quantal theory, Stevens, 1972, 1989) and allow them to delineate the boundaries of the space of possible vowels. For example, the vowel [i] determines the lower bound for the first formant, and the upper bound for the second formant, because it is produced with the narrowest possible oral cavity constriction (for a vowel) and the widest possible pharyngeal cavity

[a)]Some of the material in this article was presented at the 141st meeting of the Acoustical Society of America, Chicago, IL, 7 June 2001.
[b)]Electronic mail: francisa@purdue.edu
[c)]Electronic mail: vciocca@hkusua.hku.hk

opening. No human articulatory configuration could produce a vowel with a greater distance between $F1$ and $F2$ than that of [i]. Thus, a listener's implicit knowledge of the role of articulation in determining acoustic properties of vowels plausibly entails an understanding of the limits of possible formant configurations—the boundaries of the space of possible vowels.

This second description seems more compatible with Cowan and Morse's (1986) model, in which the boundaries of vowel trace expansion appear to function as absolute limits on the extension of the vowel memory trace. According to their model, the memory trace of a vowel is best represented as a region of probability within the vowel space. This region expands over time, representing the gradual degradation of memory acuity. As the memory of a stimulus fades, the probability of accurately recalling its formant pattern or spectral shape also decreases, or, conversely, the probability of recalling incorrect features increases. However, in this model the boundary of a memory trace cannot expand beyond the boundary of the listener's vowel space. In other words, no matter how long a time there is between the presentation of a stimulus and its recall, listeners will not have any probability of recalling the formant frequency values or spectral shape of an unpronounceable vowel. Thus, memory for tokens near a ''point'' vowel (one located close to the intersection of two boundaries) such as [i] should expand disproportionately toward the center of the listener's vowel space. The probability that listeners exposed to a prototypical [i]-like stimulus (one with very low $F1$ and very high $F2$) will remember a more [ə]-like vowel (one with lower $F2$ and higher $F1$) is much greater than that they would remember an even less [ə]-like vowel (one with an even lower $F1$ and an even higher $F2$) because such a more extreme (less neutral) vowel could not have been produced by a human vocal tract. The memory trace for an [i] cannot expand very far in the direction opposite [ə] because in that direction it is already close to the outer bounds of the listener's vowel space. Thus, the first hypothesis can be characterized as proposing that the direction of memory trace decay is a function of the structure of the perceptual space under investigation.

In contrast, Repp and Crowder (1990) argued that the effects described by Cowan and Morse (1986) are a psychophysical consequence of presentation order, and are not specific to memory for speech sounds, let alone language. In a series of experiments, Repp and Crowder (1990) found no consistent evidence that memory for vowels decays in a particular direction. For example, in their experiment 1, pairs consisting of a prototypical [ɛ] (called Pɛ) and a more [ə]-like [ɛ] (called N3ɛ) showed evidence of a decay toward [ə], in that listeners responded ''different'' less often to pairs ordered Pɛ–N3ɛ than to pairs ordered N3ɛ–Pɛ. In contrast, in pairs consisting of prototypical [i] (Pi) and a more [ə]-like [i] (N3i) showed very little evidence of a decay toward [ə], despite there being significant evidence of such effects in other experiments (Cowan and Morse, 1986; Repp *et al.*, 1979). Repp and Crowder's (1990) results suggest that the direction of vowel trace decay may depend on the particular set of stimuli used in a given experiment. According to this hypothesis, vowel contrast effects are a consequence of the gradual replacement of token-specific information with more generic information. That is, memory for exemplars is gradually replaced by information that is more representative of the category to which that exemplar most likely belongs, for example, the representation of the category prototype (cf. Hellström, 1985, and also Huttenlocher, Hedges, and Vevea, 2000). Repp and Crowder (1990) argued that Cowan and Morse's (1986) neutralization hypothesis merely represents a special case of stimuli for which the relevant generic information happens to be similar to the neutral vowel [ə]. However, Repp and Crowder (1990) conceded that their evidence was inconclusive. While some continua showed clear order of presentation effects, others did not, and the authors were unable to determine any systematic factor that might govern the observed pattern of effects. They concluded that the confusing nature of their results may be due in part to the greater acoustic complexity of speech stimuli as compared with the stimuli typically used in demonstrating effects related to the stimulus set (e.g., Braida *et al.*, 1984). Still, this second hypothesis remains plausible: The systematic distortion of memories for speech stimuli over time could result from specific experimental conditions, not as a function of listeners' knowledge of speech or language.

The goal of the present study is to more closely investigate the source of presentation order effects by examining sensitivity to small differences in fundamental frequency (f_0) across speakers of two languages. We examined the perception of stimuli differing only in f_0 by speakers of a tone language (Cantonese, where such differences can be lexically contrastive) and a nontone language (American English, where such f_0 differences are not lexically contrastive). Such stimuli have a number of advantages over vowel stimuli. First, it is possible to generate nonspeech stimuli that are acoustically quite simple (similar to those used in typical psychoacoustic studies of stimulus set-related effects), yet retain the crucial perceptual differences that cue tone category distinctions in speech stimuli. Thus, it is possible to more clearly assess the role of stimulus- or design-related factors. For example, in the examination of the present data the possibility arose that memory traces might, in some cases, decay in a unidirectional manner, regardless of category prototype location or the set of stimuli presented. According to this kind of formulation, the results presented by Cowan and Morse (1986) could be described as a decay of vowel memory traces toward the right side of the vowel space (toward a lower $F2$ value). A second advantage of using lexical tone-based stimuli is that monolingual speakers of American English do not possess a linguistically structured knowledge of pitch differences between syllables (and therefore do not have a linguistically structured ''tone space'' analogous to vowel space), while Cantonese speakers, and indeed, tone language speakers in general, do (cf. Gandour, 1981; Gandour and Harshman, 1978). By comparing the perception of small, barely suprathreshold f_0 differences by speakers of these two languages, it may be possible to determine whether contrast effects are purely a nonlinguistic consequence of the experimental stimulus set, as hypothesized by Repp and Crowder (1990), or rather a consequence of

TABLE I. Fundamental frequency values for stimuli for all experiments.

Stimulus	f_0 in Hz (mel)	Tone class
1	100.0 (150.5)	Low level
2	104.4 (156.61)	
3	108.7 (162.72)	
4	113.1 (168.83)	Mid level
5	117.5 (174.94)	Mid level
6	122.0 (181.05)	
7	126.5 (187.16)	
8	130.9 (193.27)	
9	135.5 (199.38)	
10	140.0 (205.49)	High level

some aspect of linguistic knowledge, as implied by Cowan and Morse (1986).

II. EXPERIMENT 1

The first experiment examined native Cantonese speakers' sensitivity to small suprathreshold frequency differences in synthesized Cantonese syllables ranging in f_0 along a lexical tone continuum from the frequency of a low-level tone to that of a high-level tone (see Bauer and Benedict, 1997 for a thorough discussion of Cantonese tonal phonology). In particular, this experiment was designed to investigate whether Cantonese listeners showed a difference in sensitivity due to the order of presentation of the syllables in pairs that differed by about 4 Hz.

A. Methods

1. Subjects

Fifteen native speakers of Cantonese (12 women, three men) reporting no history of speaking or hearing disability participated in this experiment. Eight were undergraduate speech pathology students in the Department of Speech and Hearing Sciences at the University of Hong Kong, and seven were students and employees from other departments. All participated in the experiment on a voluntary basis.

2. Stimuli

Stimuli for this experiment consisted of a continuum of ten 300-ms syllables synthesized with the parallel branch of a Klatt-style formant synthesizer (Klatt and Klatt, 1990) called SenSyn (Sensimetrics Corp.), implemented on a PowerMac G3. All stimuli were modeled on real Cantonese words differing only in tone (low level, corresponding to token 1, mid level, corresponding to token 4 or 5, and high level, corresponding to token 10). These words were all segmentally [ji] according to standard IPA transcription (cf. IPA, 1999). All stimuli had level f_0 contours and differed in frequency in perceptually equal steps (6.1 mel, approximately 4.4–4.5 Hz). Exact frequency values are given in Table I; each of these values was used as the fundamental frequency for the entire duration of a single stimulus syllable. Selected synthesis parameters for token 1 are given in the Appendix. Subsequent tokens differed only in f_0, as shown in Table I.

On each trial a pair of stimuli was presented with a 250-ms interstimulus interval (ISI). All pairwise combinations of syllables separated by 0 or 1 token along the continuum were presented (total of 28 pairs), including ten identical pairs (1–1, 2–2, 3–3, etc.) and 18 adjacent pairs (1–2, 2–1, 2–3, 3–2, etc.). Stimuli 1, 4, 5, and 10 were identified as satisfactory exemplars of real Cantonese words comparable to those produced by the native speaker on whose productions these stimuli were modeled. Stimulus 1 was identified as the word /ji22/ ''two'' (low level tone), stimulus 4 and 5 were identified as the word /ji33/ ''spaghetti'' (mid level tone), and stimulus 10 was identified as the word /ji55/ ''clothing'' (high level tone). Note that tones are indicated numerically, according to a commonly accepted five-point scale where 1 indicates the lowest pitch of a talker's pitch range, and 5 the highest. Two numbers are used to indicate the starting and ending pitch of the syllable. To our knowledge there is currently no published information regarding the lexical frequency or familiarity of these words in spoken Cantonese. Thus, we cannot speculate as to whether our results might have been affected by these factors. However, it may be noted that these words were selected in part because they are easily recognized and understood by children (cf. Ciocca and Lui, 2003).

3. Procedure

Participants were seated in a single-wall IAC sound booth looking through a window at the monitor of an Apple Power Macintosh 7100/AV computer located outside the sound booth. Stimuli were presented via Sennheiser HD-545 headphones at a comfortable listening level (73-dBA peak level for target words). Stimulus presentation and response collection was controlled by a Hypercard (Apple Computers, Inc.) stack running on the computer. Sounds were output through an Audiomedia II sound card at a sampling rate of 44.1 kHz. Listeners participated in two tasks, an identification task and a discrimination task. The order of the two tasks was counterbalanced across subjects (eight participants completed the discrimination task first, followed by identification, while seven completed the experiment in the reverse order). For this article, only the results of the discrimination task will be considered (Identification results are discussed in detail by Francis, Ciocca, and Ng, to appear).

The discrimination task consisted of 11 blocks, each with 28 trials. The first block was treated as familiarization and not scored, though listeners were not aware of this at the time of testing. Each trial began with the presentation of a visual warning signal on the computer screen. Subsequently, listeners heard a warning tone (an amplitude-modulated complex tone with fundamental frequency, harmonic structure, and amplitude modulation significantly different from speech) followed after 500 ms by the presentation of a single pair of syllables separated by 250-ms ISI. Following the presentation of a stimulus pair, listeners were presented with two buttons arranged horizontally on the screen, labeled *same* and *different*. Response buttons were always presented in this order. Participants were instructed to click on one of the buttons to indicate whether the syllables they heard were the same or different. After selection, followed by a brief pause, the next trial began. Order of stimulus presentation within blocks was random. Responses were collected auto-

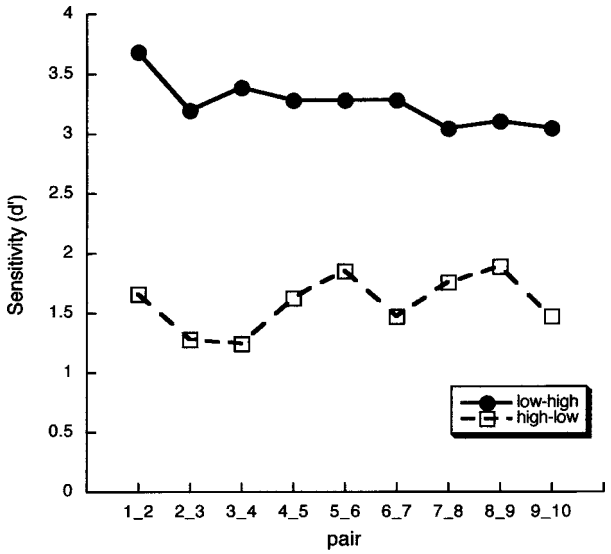


FIG. 1. Cantonese listeners' group sensitivity (d') calculated according to the roving methods (differencing model) described by Macmillan and Creelman (1991) for both orders of presentation of pairs of adjacent [ji] syllables along a continuum ranging in f_0 from 100 Hz (token 1) to 140 Hz (token 10) in perceptually equal steps (6.11 mel, approximately 4.5 Hz). Low–high order (solid line) indicates pairs in which the token with a lower f_0 is first. High–low (dotted line) indicates pairs in which the token with a higher f_0 is presented first.

matically, and scored according to whether responses were correct or not. Participants received no feedback on their responses.

B. Results

In this study we presented each stimulus pair ten times (not counting the first unscored block of trials). Group d' values based on the mean hit- and false-alarm rates of all subjects were calculated using the roving (differencing model) methods discussed by Macmillan and Creelman (1991, pp. 147–152, using Table A5.4, pp. 338–354, adapted from Kaplan, Macmillan, and Creelman, 1978). Macmillan and Creelman's Table A5.4 was generated by systematically varying the value of k (the response threshold) in the equations shown in Eq. (1) (where Φ is the normal distribution function), to arrive at a d' value for each possible H(it) and F(alse alarm) pair[1]

$$H = \Phi\left[\frac{(-k+d')}{\sqrt{2}}\right] + \left[\frac{(-k-d')}{\sqrt{2}}\right],$$
$$F = 2\Phi(-k\sqrt{2}). \tag{1}$$

Statistical analyses were carried out on the quantity (hit rate minus false alarm rate, or H-F) which served as an approximation of d' (Maddox and Estes, 1997).[2] Overall, listeners scored above chance (50%) across the continuum, ranging from 70% to 74% correct. This rate of accuracy may overestimate discrimination sensitivity because listeners appeared to be strongly biased toward responding "same" and exhibited a mean false-alarm rate of only 12% across all listeners and all stimuli.

Group d' for each stimulus pair in each order is shown in Fig. 1, where low–high indicates a pair in which the first token has a lower f_0 than the second (e.g., pair 1–2), while high–low indicates a pair in which the first token is higher in fundamental frequency than the second (e.g., pair 2–1). A two-way, repeated measures ANOVA on the difference between hit rate and false-alarm rate (H − F) showed a main effect of order of presentation (low–high mean = 0.65, high–low mean = 0.22), $F(1,14) = 19.88$, $p = 0.001$, but not of stimulus pair, $F(8,112) = 1.26$, $p = 0.27$. There was also a significant interaction between the two factors, $F(8,112) = 2.66$, $p = 0.01$. However, *post hoc* (Tukey HSD) analysis showed a significant difference ($p < 0.001$) between every one of the low–high points and every one of the high–low points, while none of the within-order pairwise comparisons reached significance ($p > 0.05$ for all). Similarly, a one-way ANOVA of the differences between the low–high and high–low scores at each pair showed a significant effect of pair, $F(8,112) = 2.66$, $p = 0.01$, but the only comparison to reach significance at the $\alpha = 0.05$ level (by Tukey HSD) involved pair 3–4, where the difference was significantly greater than that between pair 8–9.[3] Thus, the appearance of a greater overall difference at pairs 3–4 and 6–7 (near expected category boundaries) is not supported statistically.

C. Discussion

Cantonese listeners were, on the whole, more sensitive to small f_0 differences in speech stimuli when the first token in a pair had a lower f_0 (low–high order) than when the first token had a higher f_0 (high–low order). These results are consistent with the hypothesis that memory for pitch decays downward (in pitch) over time, such that pairs in the low–high order become increasingly distinct over time, while pairs in the high–low order become more similar (at least over the 250-ms ISI used in the present experiment). However, these results do not provide strong evidence for identifying the source of such a directional memory trace decay. In order to determine whether this asymmetry in sensitivity to pitch differences is related to properties of the stimuli or of the experimental procedure, as opposed to being due to the linguistic experience of the listeners, we examined the perception of listeners who had no experience making pitch-based phonological distinctions.

III. EXPERIMENT 2

Existing research on the consequences of memory trace decay suggests that biases in stimulus recall may arise from factors specific to either particular category structures (e.g., prototypes, Huttenlocher, Hedges, and Duncan, 1991), particular perceptual spaces (e.g., the geometry of vowel spaces, Cowan and Morse, 1986), or the content of particular experimental stimulus sets (Repp and Crowder, 1990). While evidence presented by Polka and Bohn (1996) suggests that language-specific vowel category prototypes are not likely to play a detectable role in determining the biased recall of vowels (at least by infants), the other two possibilities are still equally plausible. Indeed, it is even possible that tones, unlike vowels, may be influenced by language-specific category prototypes. However, the most obvious theoretical distinction is between auditory (or nonlinguistic) and linguistic

sources of contrast effects. Listeners with no experience hearing a tone language do not possess a linguistically structured ''pitch space,'' nor do they have any exclusively pitch-based phonetic categories. If these listeners show the same asymmetric pattern of discrimination as Cantonese speaking listeners, then we may conclude that this asymmetry results from the interaction of stimulus properties and human auditory capabilities. If, on the other hand, there are noticeable differences between the response patterns of the English- and Cantonese-speaking listeners, then we may conclude that these order of presentation effects are related to listeners' linguistic knowledge or experience (whether in the form of a linguistically structured perceptual space, or language-specific category prototypes).

A. Methods

1. Subjects

Nine people (five men and four women) participated in this experiment. All were native speakers of North American dialects of English. Five participants were undergraduate and graduate students and alumni from the University of Chicago, while four were newly arrived faculty members and visitors to the University of Hong Kong who had been in Hong Kong for less than a month. All of the participants reported having no knowledge of Cantonese or other tone language.

2. Stimuli

All stimuli were identical to those used in experiment 1.

3. Procedures

All procedures were identical to those described in experiment 1 including the counterbalanced participation in an additional identification experiment not reported here. However, the present experiment was run on a Macintosh iBook and stimuli were presented via Sennheiser HD-570 (three participants) or HD-580 (six participants) headphones in a quiet room. No warning tone was provided prior to stimulus presentation.[4] Stimuli were played at a comfortable listening level (approximately 75-dBA peak level for target words).

B. Results

The mean percent-correct score for American English-speaking participants ranged from 66% to 76% across the continuum, similar to Cantonese listeners. Again, American English listeners were strongly biased toward ''same'' responses, with a false-alarm rate of just 6%. As in experiment 1, group d' was calculated for each stimulus pair in each order (Macmillan and Creelman, 1991, pp. 147–151), shown in Fig. 2. For statistical comparison of the two speaker groups, both Cantonese and American English listeners' sensitivity (d') to each pair regardless of order of presentation was calculated (based on hit rate and false-alarm rate for each pair ignoring differences in order of presentation). American English listeners' mean d' across the continuum was 2.60, compared with 2.53 for Cantonese speakers, and this difference was not significant, $t(22)=0.11$, $p=0.92$. For the American listeners, a two-way repeated measures

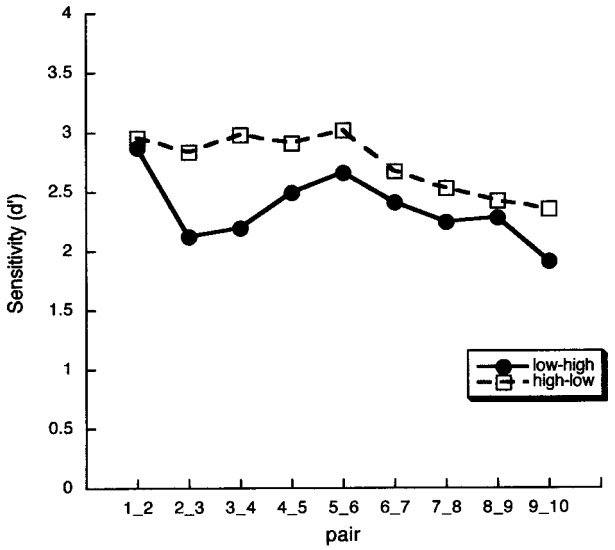


FIG. 2. American English listeners' group sensitivity (d') (calculated as for Fig. 1) for both orders of presentation of pairs of adjacent [ji] syllables along a continuum ranging in f_0 from 100 Hz (token 1) to 140 Hz (token 10) in perceptually equal steps (6.11 mel, approximately 4.5 Hz). Low–high order (solid line) indicates pairs in which the token with a lower f_0 is first. High–low (dotted line) indicates pairs in which the token with a higher f_0 is presented first.

ANOVA [calculated using (H-F) values for each order separately], showed no effect of order of presentation, $F(1,8)=3.13$, $p=0.11$, or stimulus pair, $F(8,64)=1.72$, $p=0.11$, and no interaction between the two factors, $F(8,64)=1.17$, $p=0.33$.

C. Discussion

The results of experiment 2 suggest that monolingual American English-speaking listeners are as sensitive as native speakers of Cantonese to small (subcategorical) differences in fundamental frequency. However, unlike Cantonese listeners, American listeners showed no evidence of a contrast effect in pitch discrimination. That is, there is no evidence that American English speakers are differentially sensitive to pitch differences depending on the order of presentation (low–high versus high–low). This pattern of results, when contrasted with the results of experiment 1, suggests that the source of contrast effects in Cantonese speakers' perception of tone is language specific. In this case, order-of-presentation effects are not a purely psychophysical consequence of a particular set of stimuli or experimental procedure. Experience with perceiving and speaking Cantonese apparently leads to differences in the way listeners store and/or retrieve memory traces of the pitch of auditory stimuli as compared with listeners without such experience.

The observation that speaking a tone language affects pitch perception is not new. Stagray and Downs (1993) demonstrated that speakers of a tone language (Mandarin Chinese) were less sensitive to differences between pure tones around 1000 Hz than were speakers of a nontone language (English). Stagray and Downs (1993) argued that their results can be best accounted for in terms of the categorical perception of tones. Because tone language speakers are used to making categorical decisions based on pitch, Mandarin

speakers exhibit the decreased within-category sensitivity characteristic of categorical perception of segments (cf. Macmillan, 1987 for discussion of categorical perception in signal detection-theoretic terms). However, Stagray and Down's (1993) results go beyond the usual claims of studies of categorical perception, in that they suggest that experience categorizing speech can affect sensitivity to differences between nonspeech sounds. One implication of this claim is that some kinds of speech experience may affect the function of basic auditory processes. Similar claims that linguistic experience can affect ''preattentive'' aspects of auditory processing have recently been advanced (Allen, Kraus, and Bradlow, 2000; Sharma and Dorman, 2000; Trembley *et al.*, 1997).

IV. EXPERIMENT 3

One way to demonstrate that experience with a tone language can affect basic (pitch-processing) properties of the auditory system would be to show a discrepancy in contrast effects between Cantonese and English speakers using stimuli that are not speech-like. Stagray and Downs (1993) argued that categorical perception of lexical tones was reflected in their listeners' performance on a pure-tone pitch discrimination task. If we were to observe pitch contrast effects in the processing of *nonspeech* sounds by Cantonese speakers, and if we fail to observe these contrast effects in an English-speaking population, then we may conclude that these contrast effects must result from differences in basic auditory processes related to differences in linguistic experience. This experiment was designed to compare the performance of Cantonese and English listeners on a nonspeech task equivalent to the first and second experiments.

A. Methods

1. Subjects

Two groups of listeners participated in this experiment, one Cantonese speaking, the other American English speaking. The first group consisted of nine female native speakers of Cantonese from the University of Hong Kong community, none of whom had participated in experiment 1. The second group consisted of seven native speakers of American English (four men, three women), all undergraduate or graduate students at the University of Chicago, of whom three had participated in experiment 2 two days prior to the present experiment. All participants reported normal hearing, and one Cantonese participant reported having perfect pitch.

2. Stimuli

Stimuli consisted of ten complex tones modeled on the synthetic speech stimuli used in experiments 1 and 2 and synthesized using the PowerSynthesiser application (Russell and Darwin, 1991). All stimuli were 300 ms long, consisting of eight equal-amplitude harmonics (harmonics 1, 3, 5, 6, 7, 8, 9, and 11). Harmonics 2, 4, and 10 were omitted to make the sound clearly less speech-like. Each stimulus had amplitude rise and decay times of 5 ms. The only difference between the ten stimuli was the fundamental frequency, which varied along the identical continuum as the stimuli in the first two experiments (see Table I). Complex tones were synthesized at a sampling rate of 44.1 kHz on a Power Macintosh 7100/AV equipped with an Audiomedia II sound card.

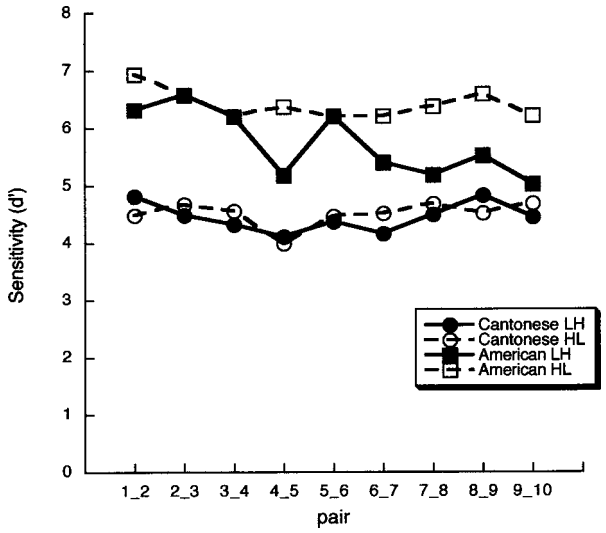


FIG. 3. Group sensitivity (d') for Cantonese listeners (circles) and American English listeners (squares) (calculated as for Figs. 1 and 2) for both orders of presentation of pairs of adjacent nonspeech tone complexes along a continuum ranging in f_0 from 100 Hz (token 1) to 140 Hz (token 10) in perceptually equal steps (6.11 mel, approximately 4.5 Hz). Low–high (solid lines) indicates pairs in which the token with a lower f_0 is first. High–low (dotted lines) indicates pairs in which the token with a higher f_0 is presented first.

3. Procedure

Procedures were identical to experiment 2 for all listeners, except that all American English listeners used Sennheiser HD-570 headphones, and all Cantonese listeners used Sennheiser HD-545 headphones and the sound presentation hardware from experiment 1.

B. Results

Group d' values were calculated for each language group based on each order of presentation, as shown in Fig. 3. In order to derive a statistical test of differences between the two groups, each listener's sensitivity (d') regardless of order of presentation was calculated for each pair (based on hit rate and false-alarm rate for each pair, ignoring differences in order of presentation) using the roving calculations described by Macmillan and Creelman (1991). Regardless of order of presentation, American English listeners' mean d' across the continuum was 5.59, compared with 4.56 for Cantonese speakers. Note that these measures of mean sensitivity were noticeably higher than those measured for speech stimuli in experiments 1 and 2 (2.60 for American listeners in experiment 2, and 2.53 for Cantonese listeners in experiment 1). A three-way mixed factorial ANOVA (between group factor of language, and repeated measures factors of pair and order) was calculated using H–F, the hit rate minus false-alarm rate statistic used in the previous two experiments. There was a significant main effect of language, $F(1,14)=6.75$, $p=0.02$, showing that American English lis-

teners were more sensitive to f_0 differences than Cantonese listeners. No other main effects or interactions were significant.

C. Discussion

Both English- and Cantonese-speaking listeners in experiment 3 were considerably more sensitive to frequency differences between these nonspeech complex tones as compared with the sensitivity of comparable groups of listeners to equivalent frequency differences in synthesized speech stimuli. This may reflect differences in the complexity of the stimuli. However, Flanagan and Saslow (1958) found difference limens (DLs) for f_0 differences between synthetic vowels (about 0.63 Hz) to be slightly *smaller* than those identified by Harris (1952) for pure tones (about 0.75 Hz) (see Klatt, 1973, for discussion).[5] In other words, increasing the acoustic complexity of stimuli does not necessarily increase the difficulty of perceiving differences in the pitch of those stimuli. Note also that DLs reported for both speech and pure tones are considerably smaller than the differences between stimuli used in the present experiments. Aside from this overall greater sensitivity to frequency differences between nonspeech stimuli, neither American English-speaking nor Cantonese-speaking listeners showed any discernible effect of order of presentation in their discrimination of nonspeech complex tones. This suggests that, whatever the source of the contrast effects observed in experiment 1, experience with Cantonese does not affect the perception of the pitch of complex tones in the same way that it affects pitch-based discrimination of speech sounds.

One final observation in experiment 3 remains puzzling. In experiment 3, Cantonese listeners exhibited an overall lower sensitivity to pitch differences (in either order) when compared with that of American English listeners. These results show some support for observations made by Stagray and Downs (1993), who found that speakers of Mandarin Chinese (also a lexical tone language, but with a different tonal inventory from Cantonese) showed significantly larger frequency difference limens than did native speakers of English on a pure-tone discrimination task. Stagray and Downs (1983) attributed this difference to Mandarin speakers' categorical perception of tone. They argued that the frequency differences between the stimulus pairs used in their experiment were always well within a single category, within which acuity should be lower than across category boundaries according to standard theories of categorical perception [e.g., Liberman *et al.* (1957)]. However, there are two problems with using categorical perception to account for the results of experiment 3. First, although there was a significant difference between the sensitivities of the two language groups on the nonspeech complex tone discrimination task (experiment 3), the difference between their discrimination sensitivity on the speech task (experiment 1 versus experiment 2) was not significant. If Cantonese speakers' poorer discrimination of nonspeech tones was due to their greater experience with making pitch-based category judgments of speech stimuli, then there should be a similar, if not greater, difference in sensitivity between the two language groups when judging speech-like stimuli. However, the two groups were not significantly different in their overall sensitivity to f_0 differences between speech stimuli. They only showed a difference in sensitivity to nonspeech sounds. Second, the frequency range employed in these experiments was selected to encompass the three level tone categories of Cantonese. In particular, token pairs 3–4 and 6–7 explicitly span the category boundaries. Although the results of experiment 3 contribute to the mounting evidence that tone language speakers tend to be less sensitive than English speakers to fundamental frequency differences between nonspeech sounds (Stagray and Downs, 1993; Tanner and Rivette, 1964, though see Burns and Sampat, 1980 for a counterexample to this trend), there appears thus far to be little evidence that this tendency is related to the categorical perception of tone as typically specified.

V. GENERAL DISCUSSION

The results of experiment 1 suggest that Cantonese listeners' ability to discriminate between level fundamental frequency contours is strongly influenced by the order in which pairs of stimuli are presented. When the first token in a pair is lower in f_0 than the second (low–high order), listeners are considerably more sensitive to the difference than when the order of pairs is reversed (high–low order). One way to account for this difference is in terms of a gradual decay of the memory trace of the initial token, such that it is recalled as having a lower pitch for the purposes of comparison with the later-occurring token. Although the results of experiment 1 suggest that such a memory trace decay would be directional (toward lower values), we found no strong evidence to suggest that stimulus set related properties could have contributed to the appearance of these effects. If the direction of memory trace decay were the result of properties of the stimulus set, we would expect the directionality to be symmetrical, either with relation to the edges of the continuum or with respect to some more centrally located region along the continuum. With respect to the role of edges in the directionality of memory trace decay, Macmillan, Braida, and Goldberg (1987) characterize Berliner *et al.*'s (1977) *bias edge effect* in terms of the boundaries (edges) of the stimulus continuum. Berliner *et al.* (1977) found that when listeners were presented with two tokens of overall low intensity (close to the low-intensity end of the continuum) they tended to hear the first token as louder than the second. But, when the same intensity difference was presented using high-intensity stimuli (at the high-intensity end of the continuum), listeners tended to hear the first token as quieter than the second. Thus, the location of the stimulus pair along the continuum affected the direction of the bias—the first token of a quiet pair was heard as louder while the first token of a loud pair was heard as quieter, suggesting that in both cases the memory trace of the first token decayed toward a more intermediate (centrally located) value along the stimulus continuum.

In the present case, we do observe something like a bias edge effect at the right (higher frequency) end of the continuum, in that the first token in a higher-frequency pair (e.g., 8–9 or 9–8) is generally heard as having a higher pitch (such that 8–9 is treated more like 9–9, a ''same'' pair by being

only poorly discriminable, while 9–8 is still quite easily discriminated). This pattern can clearly be characterized as a decay of the memory for pitch of the first token toward a lower (more central) value along the continuum. However, there is no corresponding upward decay of the memory of the pitch of the first token in a lower frequency pair (e.g., 2–3). In this case we would expect to observe a tendency for the first token to be remembered as *higher* in pitch (e.g., for pair 2–3 to be poorly discriminated), but this is not what was observed. Instead, we see a tendency for the first token to be treated as *always* having a lower pitch than the second, regardless of where along the continuum the two tokens are located. Similarly, there does not seem to be any location along the continuum toward which, or away from which, memory traces seem to decay. The trend is always in a downward direction across the entire continuum. Thus, although there is clear evidence that Cantonese listeners' contrast effects can be described in terms of a general decay of memory for pitch toward lower values, there is no explicit evidence that this effect is due to properties of the stimulus set itself.

The results of experiment 2 further support the hypothesis that asymmetric discrimination of f_0 differences in synthetic speech stimuli by Cantonese listeners is not due to some property of the stimulus set, but rather is related to listeners' knowledge of Cantonese, a lexical tone language. In experiment 2, American English-speaking listeners with no experience with any lexical tone language showed no evidence of any order of presentation effects when tested with the same stimuli and procedures as in experiment 1. It is important to note that the American listeners of experiment 2 were (in both orders of presentation) about as sensitive to the f_0 differences between these stimuli as were the Cantonese listeners of experiment 1. The results of experiment 1 and 2, taken together, suggest that the contrast effects shown by the Cantonese listeners are a consequence of their knowledge of Cantonese.

The results of experiment 3 suggest that, whatever the specific mechanism that causes contrast effects in Cantonese listeners' perception of spoken pitch, it does not appear to have affected their ability to discriminate the pitch of nonspeech sounds. Cantonese listeners, like American English listeners, showed no asymmetry in sensitivity to f_0 differences between nonspeech stimuli that were identical in f_0 to the speech stimuli used in experiments 1 and 2. These results provide further support that the order of presentation effects demonstrated by Cantonese listeners are a consequence of their linguistic processing of speech stimuli.

A. The role of language experience in memory trace decay

One way to account for the difference in the effects of order of presentation for speakers of tone vs nontone languages might be to conclude that American listeners, lacking experience with pitch-based lexical distinctions, do not have a mental representation of a ''tone space.'' As a result, American English listeners' memories for words do not decay in a manner that is affected by the boundaries of such a space. Cantonese listeners, on the other hand, can be described as basing their perceptual judgments on relative relations between the mental representations of tokens in a tone space (cf. Gandour, 1981). Following the model proposed by Cowan and Morse (1986), we might expect the boundaries of Cantonese listeners' tone space to cause an asymmetric expansion of the memory trace for the pitch of the earlier-presented syllable away from the nearer boundary of the tone space (in the same way that memory for [i]-like stimuli is proposed to expand disproportionately away from the high-front corner of a listener's vowel space). There is, however, one significant problem with such an account. The stimuli used in experiments 1 and 2 range in frequency from a prototypical low-level tone to a prototypical high-level tone, encompassing the vast majority of the normal spoken frequency range of the speaker on whose productions they are modeled. Any explanation of order of presentation effects based on the influence of the boundaries of tone space on memory traces would predict opposing effects at either end of a frequency continuum spanning that space. That is, tokens at the high end of the continuum should decay toward a lower pitch as their memory trace, located near the top end of the space, cannot expand very far in a higher direction. Conversely, tokens at the low end of the continuum should decay toward a higher pitch. Contrary to this hypothesis, however, all tokens along the continuum, from lowest to highest, show an asymmetry in discriminability between the low–high and high–low orders, such that a decay-based account would have to conclude that the memory trace of every token appears to decay toward a lower pitch value.

Another possibility is that listeners' memory for pitch might decay, or be biased, toward tone-category prototypes (or away from category boundaries) as suggested by Huttenlocher and her colleagues (Huttenlocher *et al.*, 1991; Huttenlocher *et al.*, 2000). However, on a tone identification task using the same stimuli (Francis *et al.*, to appear) Cantonese listeners showed category boundaries between tokens 3 and 4 (between the categories low level and midlevel) and 7 and 8 (between the categories midlevel and high level).[6] Thus, if memory for pitch decayed toward category prototypes or away from category boundaries, we would expect listeners' sensitivity to pairs spanning these boundaries to be relatively high regardless of order of presentation. For example, if the 3–4 pair were presented in the low–high order (first token 3, then token 4), then the memory for 3 should decay toward the prototype for the low-level category (somewhere around token 1). Discrimination should be accurate because the perceived difference between the two syllables should increase due to category-related bias. If this pair were presented in high–low order (first token 4, then token 3), memory for token 4 should decay toward the prototype for the midlevel category (somewhere around token 5), similarly improving discrimination. However, discrimination was considerably better for the low–high order for this, and most other, pairs of tokens (including pair 7–8, the other cross-boundary pair). Thus, the results of experiment 1 suggest that neither the geometry of tonal space in general, nor listeners' language-specific phonological inventory, play a primary role in determining the directionality of the observed contrast effects. None of the results of experiment 1 provides explicit support for a memory trace decay model. There is no evi-

dence that the order of presentation effects shown in experiment 1 result from basic psychoacoustic properties of the stimulus or task. In contrast, the results of experiments 2 and 3 demonstrate that order of presentation effects in level tone perception occur only when native speakers of Cantonese are listening to speech.

B. Discrimination asymmetries and f_0 declination

The mechanism of memory trace decay could account for the present results if we arbitrarily assume that memories for pitch decay toward lower pitches. However, since there is, at present, no corroborating evidence to support this assumption, it may prove useful to investigate other mechanisms that could account for these data. One aspect of Cantonese speech production that may prove useful in this regard is the phenomenon of f_0 declination or downdrift, the gradual declination of fundamental frequency over the course of an utterance (Ohala, 1978; Pierrehumbert, 1979; Umeda, 1982; Vaissière, 1995; Vance, 1976; Wong, 1999). Downdrift has been argued to be a universal, language-independent (and even cross-species) characteristic of speech production (Hauser and Fowler, 1992; Ohala, 1978). In English, short declarative utterances without prominent focal stress, such as those typically elicited in laboratory speech experiments, tend to exhibit clear declination of f_0 over time (Umeda, 1982). Furthermore, American English listeners show evidence of compensating for downdrift in the perception of the relative pitch of early- and late-occurring syllables in sentences. For example, Pierrehumbert (1979) showed that listeners perceived a syllable occurring early in an utterance as having equal pitch with a later-occurring syllable even though the later syllable had a lower f_0. This was interpreted as evidence that listeners were able to compensate perceptually for an expected declination in pitch over the course of the utterance.

There is evidence that Cantonese speakers also exhibit f_0 declination in speech production (Vance, 1976), and some suggestion that Cantonese listeners perceptually compensate for this expected declination. For example, Wong (1999) presented listeners with a target word preceded by a context sentence in which the f_0 had been manipulated in one of two ways. The sentence was divided in half and the f_0 of either the earlier-occurring portion or the later-occurring portion of the context sentence was shifted. Results suggested that Cantonese listeners based their judgments of the tone of the target word on the pitch of more recent (later-occurring) pitch information in the sentence. For example, if the f_0 of the second half of the sentence was shifted down, listeners responded as if the target word had a high level tone. When the f_0 of the second half of the sentence was shifted upward, listeners tended to respond as if the word had a low-level tone, although in both cases the f_0 of the target world remained the same. However, when the f_0 of the second half of the sentence was shifted upward, the expectation of downdrift was violated. In this case listeners did not respond as strongly according to the more recent (second half) pitch information. That is, they made fewer than expected identifications of the target syllable as having a low-level tone. Wong (1999) argued that listeners responded less strongly in this condition because they were confused by the violation of their expectations for downdrift.

It is possible to account for the results found in experiment 1 in terms of a compensation for a learned expectation that utterances will tend to exhibit a slight declination in f_0 from beginning to end. In order to correctly identify the tone of a syllable, Cantonese listeners must be able to take into account the position of that syllable within the utterance. Syllables at the end of the sentence must be perceptually raised in pitch, otherwise they risk being identified as having a lower tone than the speaker intended. In the case of the high–low order of presentation in experiment 1, this raising of the perceived pitch results in the two tokens sounding similar or identical. In contrast, in the low–high presentation order this raising results in a heightened perception of the difference. This hypothesis is supported by the observation that, on average, Cantonese listeners were slightly more sensitive than American listeners to pairs presented in the low–high order, but less sensitive than American listeners to pairs presented in the high–low order.

One problem remains. American English listeners in the present experiment showed no evidence of an expectation that pitch should fall over the course of an utterance, although such expectations have previously been demonstrated in English listeners' judgments of the pitch of syllables in sentential context (Pierrehumbert, 1979). It is possible that American English listeners did not treat the syllable pairs used in this experiment as a single utterance. One possible reason for this is that the syllable [ji] is not an English word. However, Pierrehumbert (1979) achieved her results using nonsense sentences made up of repetitions of the syllable [ma]. A more likely explanation is based on the observation that English is a stress-timed language (Laver, 1994, p. 157), while Cantonese is typically considered to be syllable-timed (Bauer and Benedict, 1997, p. 316). Thus, American English listeners presumably expect utterances to exhibit a pattern of more or less alternating stressed (louder, higher pitched, longer) and unstressed (quieter, lower pitched, shorter) syllables.

According to this explanation, there are two reasons that speakers of a syllable-timed language would not show perceptual compensation for downdrift in the stimuli used here. First, American listeners might not have treated the two syllables in experiment 2 as a single utterance, perhaps because both syllables were equally loud. Alternatively, they may have treated it as a single utterance consisting of a single word, perhaps because the alternation in pitch suggested the presence of one stressed and one unstressed syllable. In the first case listeners might have expected a reset of the pattern of pitch declination with the start of the second utterance. The resetting of pitch declination at utterance breaks is a common phenomenon according to Pierrehumbert (1979), and might plausibly enable listeners to accurately distinguish small f_0 differences between the two syllables because both syllables are located at the start of the expected declination curve. In the second case, it is possible that American English speakers only compensate for pitch declination between *stressed* syllables. Since stressed syllables are typically sepa-

rated by at least one unstressed syllable (as in Pierrehumbert's experiments), American English listeners might retain the ability to distinguish small f_0 differences between adjacent syllables. Indeed, such an ability might contribute to the ability to distinguish between stressed and unstressed syllables. Further research on the perception of pitch, and stress, in American English utterances would be necessary to distinguish between these two possibilities.

ACKNOWLEDGMENTS

Some of the results presented in experiment 1 were first noted in a dissertation submitted by Brenda Ng in partial fulfillment of the requirements for the B.Sc. degree in Speech and Hearing Sciences at the University of Hong Kong. We are grateful to Howard Nusbaum and Kimberly Fenn at the University of Chicago for their assistance with running experiments 2 and 3, and to Neil Macmillan for insightful discussions concerning signal detection theory. Some of the results discussed here were first presented at the 141st meeting of the Acoustical Society of America, Chicago, IL, 7 June 2001. This research was conducted while the first author was a postdoctoral fellow at the University of Hong Kong, in the Department of Speech and Hearing Sciences, and was supported in part by the Hong Kong Research Grants Council, HKU 7193/00H, to Valter Ciocca.

APPENDIX: SYNTHESIS PARAMETERS

Stimulus synthesis used parameters measured at 1-ms intervals from a natural speech sample. After initial synthesis, parameter values were subsequently smoothed by hand, resulting in roughly linear contours for major frequency and amplitude parameter values. Stimuli were synthesized using an update interval of 5 ms. Parameter AV (amplitude of voicing) rose from a value of 50 to 60 dB over the first 210 ms of the utterance, and then declined to a value of 42 dB at the end of the utterance. $F1$ rose from 360 to 368 Hz over the first 100 ms and remained level for the remainder of the utterance. $F2$ rose from 2308 to 2392 Hz over the first 100 ms, and then fell to 2193 Hz by the end of the utterance. $F3$ began at 3712 Hz and fell to 3574 Hz in the first 100 ms, then fell more gradually to 2929 Hz at the end of the utterance. $F4$ began at 4126 Hz, and fell to 3826 Hz by 195 ms, then to 3620 Hz at the end of the utterance. $F5$ began at 4586 Hz, and fell to 4279 Hz by the end of the utterance. A1V (amplitude of $F1$) began at 55 dB and rose to 59 dB by 170 ms, remained level until 220 ms, then fell to 56 dB at 260 ms, and then fell more sharply to 42 dB by the end of the utterance. AV2, AV3, and AV4 all began at 46 dB. AV2 rose to 57 dB at 140 ms, then fell to 56 dB at 160 ms, remained level to 205 ms, fell to 55 dB by 210 ms, remained level until 275 ms, then fell to 42 dB by the end of the utterance. AV3 rose to 57 dB at 145 ms, then starting at 260 ms fell to 42 dB by the end of the utterance. AV4 rose to 55 dB at 145 ms, then remained level until 260 ms, at which point it fell to 42 dB by the end of the utterance.

[1]In this experiment we are interested in separately analyzing pairs of stimuli with different orders of presentation. Therefore, there are half as many trials for each "different" pair *in each order* (10) as there are "same" trials (20). Because there is a maximum of ten possible hits for a given order of presentation, but there are up to 20 possible false alarms, correcting for perfect scores according to the methods proposed by Macmillan and Creelman (1991, p. 10) would lead to different z scores for a hit rate of 1.0 and a false-alarm rate of 1.0 (and similarly for scores of 0.0). This is not a problem for analyzing group data, since there were no perfect scores on any pair. However, some individual subjects did score either perfect hit rates (1.0) or perfect false-alarm rates (0.0) on one or another stimulus pair. Therefore, individual d' scores were not calculated.

[2]Because this is not a commonly used statistic, all analysesusing the H−F statistic were also repeated using arcsine-transformed percent correct [P(C)] values. Unless otherwise noted, tests reported as significant based on the H−F score were also found to be significant using the P(C) score, while results reported as not significant were also found not to be significant using the P(C) score, although exact F- and p-values did differ between tests using the two scores.

[3]Note that analysis of arcsine-transformed P(C) data showed a significant difference between pairs 1–2 and 8–9 and 3–4 and 8–9.

[4]Although this difference in procedure could conceivably have contributed to the pattern of results observed here, we have also subsequently conducted additional tests with Cantonese listeners and without a warning tone and have observed results qualitatively similar to those reported in experiment 1.

[5]Flanagan and Saslow (1958) used a set of synthetic vowels (/ɑ/) with a standard f_0 of 120 Hz and 70 dB SPL and found a median of 0.63 Hz, while Harris (1952) used pure tones with a standard reference tone of 125 Hz at 30 dB SL and found a median of 0.74 Hz.

[6]Although the experiments conducted by Francis *et al.* (to appear) provided clear evidence for the presence of category boundaries using an identification paradigm, the discrimination testing results they report showed no evidence for the categorical perception of level tones in Cantonese. The question of whether lexical tones are perceived categorically is discussed in detail by Francis *et al.* (to appear).

Allen, J., Kraus, N., and Bradlow, A. (**2000**). "Neural representation of consciously imperceptible speech sound differences," Percept. Psychophys. **62**(7), 1383–1393.

Bauer, R. S., and Benedict, P. K. (**1997**). *Modern Cantonese Phonology* (Mouton de Gruyter, Berlin).

Berliner, J. E., Durlach, N. I., and Braida, L. D. (**1977**). "Intensity perception. VII. Further data on roving level discrimination and the resolution of bias edge effects," J. Acoust. Soc. Am. **61**, 1577–1585.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context-coding," J. Acoust. Soc. Am. **76**, 722–731.

Burns, E. M., and Sampat, K. S. (**1980**). "A note on possible culture-bound effects in frequency discrimination," J. Acoust. Soc. Am. **68**, 1886–1888.

Ciocca, V., and Lui, J. Y. K. (**2003**). "The development of the perception of Cantonese lexical tones," J. Multiling. Commun. Disord. **1**, 141–147.

Cowan, N., and Morse, P. A. (**1986**). "The use of auditory and phonetic memory in vowel discrimination," J. Acoust. Soc. Am. **79**, 500–507.

Flanagan, J. L., and Saslow, M. G. (**1958**). "Pitch discrimination for synthetic vowels," J. Acoust. Soc. Am. **30**, 435–442.

Francis, A. L., Ciocca, V., and Ng, B. C. K. (to appear). "On the (non)categorical perception of lexical tones," Percept. Psychophys.

Gandour, J. T., and Harshman, R. (**1978**). "Crosslanguage differences in tone perception: A multidimensional scaling investigation," Lang. Speech **21**, 1–33.

Gandour, J. T. (**1981**). "Perceptual dimensions of tone: Evidence from Cantonese," J. Chin. Ling. **9**, 20–36.

Harris, J. D. (**1952**). "Pitch discrimination," J. Acoust. Soc. Am. **24**, 750–755.

Hauser, M. D., and Fowler, C. A. (**1992**). "Fundamental frequency declination is not unique to human speech: Evidence from nonhuman primates," J. Acoust. Soc. Am. **91**, 363–369.

Hellström, A. (**1985**). "The time-order error and its relatives: Mirrors of cognitive processes in comparing," Psychol. Bull. **97**, 35–61.

Huttenlocher, J., Hedges, L. V., and Duncan, S. (**1991**). "Categories and particulars: Prototype effects in estimating spatial location," Psychol. Rev. **98**(3), 352–376.

Huttenlocher, J., Hedges, L. V., and Vevea, J. L. (**2000**). "Why do categories affect stimulus judgment?," J. Exp. Psychol. Gen. **129**(2), 220–241.

IPA (**1999**). *Handbook of the International Phonetic Association* (Cambridge University Press, Cambridge).

Klatt, D. H. (**1973**). ''Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception,'' J. Acoust. Soc. Am. **53**(1), 8–16.
Klatt, D. H., and Klatt, L. C. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,' J. Acoust. Soc. Am. **87**, 820–857.
Ladefoged, P. (**2001**). *A Course in Phonetics*, 4th ed. (Thompson Learning, Australia), pp. 78–79.
Laver, J. (**1994**). *Principles of Phonetics* (Cambridge University Press, Cambridge), p. 157.
Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). ''The discrimination of speech sounds within and across phoneme boundaries,'' J. Exp. Psychol. **54**, 358–368.
Lieberman, P., and Blumstein, S. E. (**1988**). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge University Press, Cambridge).
Macmillan, N. A. (**1987**). ''Beyond the categorical/continuous distinction: A psychophysical approach to processing modes,'' in *Categorical Perception*, edited by S. Harnad (Cambridge University Press, Cambridge), pp. 53–85.
Macmillan, N. A., Braida, L. D., and Goldberg, R. F. (**1987**). ''Central and peripheral processes in the perception of speech and nonspeech sounds,'' in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Martinus Nijhoff, Dordrecht, The Netherlands), pp. 28–45.
Macmillan N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).
Maddox, W. T., and Estes, W. K. (**1997**). ''Direct and indirect stimulus-frequency effects in recognition,'' J. Exp. Psychol. Learn. Mem. Cogn. **23**(3), 539–559.
Ohala, J. J. (**1978**). ''Production of tone,'' in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academic, New York), pp. 5–39.
Pierrehumbert, J. A. (**1979**). ''The perception of fundamental frequency declination,'' J. Acoust. Soc. Am. **66**, 363–369.
Polka, L., and Bohn, O.-S. (**1996**). ''A cross-language comparison of vowel perception in English-learning and German-learning infants,'' J. Acoust. Soc. Am. **100**(1), 577–592.
Repp, B. H., and Crowder, R. G. (**1990**). ''Stimulus order effects in vowel discrimination,'' J. Acoust. Soc. Am. **88**(5), 2080–2090.
Repp, B. H., Healy, A. F., and Crowder, R. G. (**1979**). ''Categories and context in the perception of isolated steady-state vowels,'' J. Exp. Psychol. Hum. Percept. Perform. **5**(1), 129–145.
Russell, P., and Darwin, C. J. (**1991**). ''Real-time synthesis of complex sounds on a Mac II with 56001 DSP chip,'' Br. J. Audiol. **25**, 59–60.
Sharma, A., and Dorman, M. F. (**2000**). ''Neurophysiologic correlates of cross-language phonetic perception,'' J. Acoust. Soc. Am. **107**(5), 2697–2703.
Stagray, J. R., and Downs, D. (**1993**). ''Differential sensitivity for frequency among speakers of a tone and a non-tone language,'' J. Chin. Ling. **21**, 144–163.
Stevens, K. N. (**1972**). ''The quantal nature of speech: Evidence from articulatory-acoustic data,'' in *Human Communication: A Unified View*, edited by E. E. David, Jr. and P. B. Denes (McGraw-Hill, New York), pp. 51–66.
Stevens, K. N. (**1989**). ''On the quantal nature of speech,'' J. Phonetics **17**, 3–45.
Tanner, W. P., and Rivette, G. L. (**1964**). ''Experimental study of 'tone deafness,' '' J. Acoust. Soc. Am. **36**, 1465–1467.
Tremblay, K., Kraus, N., Carrell, T. D., and McGee, T. (**1997**). ''Central auditory system plasticity: Generalization to novel stimuli following listening training,'' J. Acoust. Soc. Am. **102**(6), 3762–3773.
Umeda, N. (**1982**). '''F_0 declination' is situation dependent,'' J. Phonetics **10**, 279–290.
Vaissiére, J. (**1995**). ''Phonetic explanations for cross-linguistic prosodic similarities,'' Phonetica **52**, 123–130.
Vance, T. J. (**1976**). ''An experimental investigation of tone and intonation in Cantonese,'' Phonetica **33**, 368–392.
Wong, P. C. M. (**1999**). ''The effect of downdrift in the production and perception of Cantonese level tone,'' in Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco, Vol. 3, pp. 2395–2398.

A theoretical study of the vibration and acoustics of ancient Chinese bells

M. Jing[a)]
Institute of Mechanics, School of Civil Engineering and Architecture, Northern Jiaotong University, 100044 Beijing, People's Republic of China

(Received 4 April 2002; revised 6 June 2003; accepted 23 June 2003)

In this paper, the acoustics of an ancient Chinese bell, which was made some 3000 years B. C., is studied theoretically. In ancient times, a set of the bells was used as a musical instrument. Unlike a western church bell and an ancient Indian bell, an ancient Chinese bell has two interesting acoustics. First, two tones can be heard separately as the bell is struck at two special points. The interval between the two pitches is always a minor or major third. Second, tones of the bell attenuate quickly, which is necessary for a musical instrument. So, an ancient Chinese bell is sometimes called a two-tone bell or a music bell. Although a three-dimensional model should be used to simulate the acoustics of the bell, a simplified model proposed in this paper does give some insight. Based on the lens-shaped cross section of an ancient Chinese bell, two tones of an ancient Chinese bell can be simulated by the vibration of a double-circular arch and the quick attenuation of tones can be simulated by acoustics of a cylinder with the lens-shaped cross section like a double-circular arch. Numerical results on the vibration and acoustics of the models are presented. [DOI: 10.1121/1.1600727]

PACS numbers: 43.75.Kk, 43.20.Hq [NHF]

I. INTRODUCTION

In the 1950s and 1970s, two complete sets of ancient Chinese bells were discovered in *Henan* province and *Hubei* province in China, respectively. A sketch of the bell is shown in Fig. 1.

These ancient bells have two interesting acoustics. First, two tones can be heard separately as two special points on the bell are struck. The two special points, inscribed on the bell as shown in Fig. 1, are called *SUI* and *GU* (*SUI* and *GU* are pronunciation of two words in ancient Chinese language), respectively. The *SUI* pitch is a minor or major third lower than *GU* pitch. Second, the tones of the bell attenuate quickly, without prolonged echo. This is called short duration of sound. Sinyang[1] described in detail the acoustics of the bell.

Unlike a western church bell or an ancient Indian bell, the cross section $a-a$ of an ancient Chinese bell is lens-shaped or eye-shaped as shown in Fig. 1. Why is the cross section not round? The earliest record on short duration of sound of a Chinese bell can be found in an ancient Chinese book *Mengxi Bitan* (Dream Stream Essays) written by *Shen Kuo* (an ancient Chinese statesman and scientist, 1030–1095, Northern Song Dynasty). Through observation, *Shen Kuo* noted that the tone of a round bell has more prolonged echo than that of an oblate bell, and that an oblate bell is suitable for a musical instrument. In the view of modern science, the acoustics of the ancient Chinese bell is determined by its special structure.

Sinyan[1] studied an ancient Chinese bell experimentally. Rossing *et al.*[2] investigated vibration modes of a Chou Dynasty bell by means of holographic interferometry and by scanning the sound field near the bell.

Theoretically, a three-dimensional model should be used to simulate acoustics of the bell. To the best of our knowledge, there is no such report on a theoretical study of the bell. Because a three-dimensional model is very complex, a simplified model is proposed in this paper to investigate the acoustics of the bell. Based on the lens-shaped section of the bell, it is supposed that free vibration of cross section of an ancient Chinese bell can be simulated by the free vibration of a double-circular arch as shown in Fig. 2. Interests in the natural vibration of arch can be traced back to the 19th century. Chidamparam[3] gave a survey on vibration of an arch. In this paper, the effects of shear deformation and rotary inertia in vibration are neglected. By extensional theory of planar

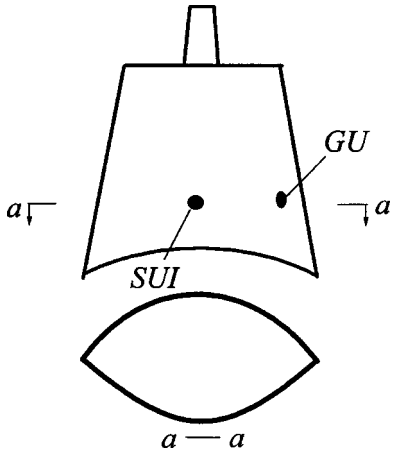


FIG. 1. Sketch of an ancient Chinese bell. There are two striking points *SUI* and *GU* corresponding to two tones. Tones of the bell attenuate quickly, without prolonged echo. The cross section $a-a$ of the bell is lens-shaped, which is similar to a double-circular arch.

a)Electronic mail: jinmingjinming@hotmail.com

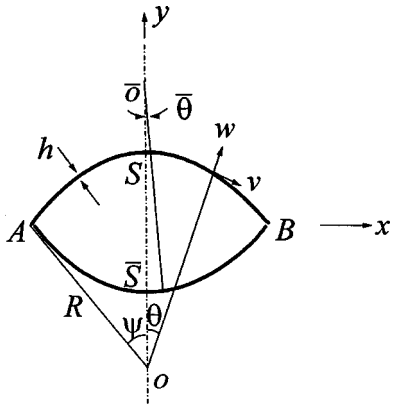


FIG. 2. A double-circular arch is made of two circular arches ASB and $\overline{\text{ASB}}$ with radius R. h denotes thickness of each arch, ψ denotes a half of central angle of each arch, and θ denotes coordinate along the arch ASB. w denotes the normal displacement of centroidal axis, while v denotes the tangential component. Symbols with an upper bar "‾" denote parameters on the arch ASB.

curved beam, the basic vibration equations for a circular arch have been obtained by Chidamparam[3] and Soedel.[4] Exact solution of the basic vibration equations, which Chidamparam[5] studied, is used to analyze the vibration of the double-circular arch. Furthermore, it is supposed that short duration of sound can be simulated by acoustics of a vibrating cylinder with the lens-shaped cross section like a double-circular arch. According to Morse,[6] we can get the total power radiated in an unit time from the vibrating cylinder, which is the source of sound field. Then, we get attenuation exponential of sound to explain the short duration of sound.

II. VIBRATION OF A DOUBLE-CIRCULAR ARCH

A. Exact solution of natural mode on planar circular arch

Because the cross section of an ancient Chinese bell is lens-shaped or eye-shaped, we suppose that the shape of the cross section is a double-circular arch which is made of a circular arch ASB and a circular ASB, each with radius R as shown in Fig. 2. Here, h denotes thickness of each arch, ψ denotes a half of the central angle of each arch, ρ denotes mass density, E denotes Young's modulus, and θ denotes angular coordinate along arch ASB, w denotes the normal displacement along centroidal axis, while v denotes tangential displacement. In Fig. 2, symbols with an upper bar "‾" denote parameters on arch ASB.

In extensional theory, the dynamic equations for a circular arch undergoing planar vibration are Eqs. (12a), (12b) in Chidamparam[3] as

$$(1+k^2)\frac{\partial^2 v}{\partial\theta^2}+\left(-k^2\frac{\partial^3}{\partial\theta^3}+\frac{\partial}{\partial\theta}\right)w=\frac{\rho R^2}{E}\frac{\partial^2 v}{\partial t^2}$$
$$\left(-k^2\frac{\partial^3}{\partial\theta^3}+\frac{\partial}{\partial\theta}\right)v+\left(k^2\frac{\partial^4}{\partial\theta^4}+1\right)w=-\frac{\rho R^2}{E}\frac{\partial^2 w}{\partial t^2}, \tag{1}$$

where $k^2=h^2/12R^2$.

For the circular arch ASB, let

$$v=V(\theta)e^{j\omega t},\quad w=W(\theta)e^{j\omega t}, \tag{2}$$

where V and W are the amplitudes of tangential and normal vibration modes, respectively, and ω is the natural frequency.

Substituting (2) into (1), we have

$$\left[(1+k^2)\frac{d^2}{d\theta^2}+\frac{\rho R^2}{E}\omega^2\right]V+\left(-k^2\frac{d^3}{d\theta^3}+\frac{d}{d\theta}\right)W=0$$
$$\left(-k^2\frac{d^3}{d\theta^3}+\frac{d}{d\theta}\right)V+\left(k^2\frac{d^4}{d\theta^4}+1-\frac{\rho R^2}{E}\omega^2\right)W=0. \tag{3}$$

From (3)

$$\Delta V=0, \tag{4}$$

$$\Delta W=0, \tag{5}$$

where Δ is an operator

$$\Delta=\left[(1+k^2)\frac{d^2}{d\theta^2}+\frac{\rho R^2}{E}\omega^2\right]\left(k^2\frac{d^4}{d\theta^4}+1-\frac{\rho R^2}{E}\omega^2\right)-\left(-k^2\frac{d^3}{d\theta^3}+\frac{d}{d\theta}\right)^2. \tag{6}$$

Letting $W(\theta)=e^{\lambda\theta}$, we have the character equation of Eqs. (4) and (5)

$$\lambda^6+a_2\lambda^4+a_1\lambda^2+a_0=0, \tag{7}$$

where

$$a_0=\frac{1}{k^2}\frac{\rho R^2}{E}\omega^2\left(1-\frac{\rho R^2}{E}\omega^2\right),$$
$$a_1=1-\left(1+\frac{1}{k^2}\right)\frac{\rho R^2}{E}\omega^2, \tag{8}$$
$$a_2=2+\frac{\rho R^2}{E}\omega^2.$$

Equation (7) is a cubic equation in λ^2. So, we have a set of roots in (7) for different ω. For a different set of roots in (7), we have the exact solution of (5) as

$$W(\theta)=\mathbf{c}^T\mathbf{f}, \tag{9}$$

where $\mathbf{c}_{6\times1}$ is a vector with six unknown constants for arch ASB and $\mathbf{f}_{6\times1}$ is solution vector in (5) for a set of roots of (7). Similarly, we have the exact solution V in (4)

$$V(\theta)=\tilde{\mathbf{c}}^T\mathbf{f}, \tag{10}$$

where $\tilde{\mathbf{c}}_{6\times1}$ is a vector with six other unknown constants for arch ASB and $\mathbf{f}_{6\times1}$ is the same solution vector as that in Eq. (9).

Let

$$\frac{d\mathbf{f}}{d\theta}=\mathbf{L}\mathbf{f}, \tag{11}$$

where $\mathbf{L}_{6\times6}$ is a matrix. Substituting (9), (10), (11) into the first equation in (3), we have the relation between $\mathbf{c}$ and $\tilde{\mathbf{c}}$

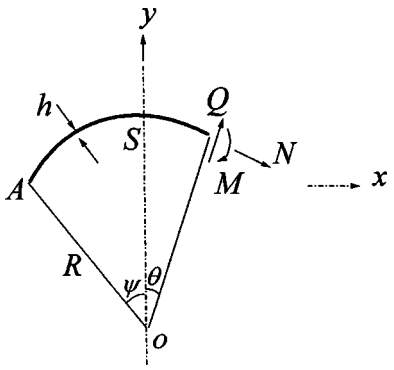


FIG. 3. *N*, *M*, and *Q* denote mode of internal forces: axial tension, moment, and shear force, respectively, of the arch ASB.

$$\tilde{\mathbf{c}}^T = \mathbf{c}^T \mathbf{G}, \tag{12}$$

where $\mathbf{G}_{6\times 6}$ is a matrix

$$\mathbf{G} = (k^2\mathbf{L}^3 - \mathbf{L})\left((k^2+1)\mathbf{L}^2 + \frac{\rho R^2}{E}\omega^2\mathbf{I}\right)^{-1}, \tag{13}$$

where $\mathbf{I}_{6\times 6}$ is unit matrix. Substituting (12) into (10), we have

$$V(\theta) = \mathbf{c}^T\mathbf{G}\mathbf{f}. \tag{14}$$

Equations (9) and (14) are the exact solution of the normal and tangential component of the natural mode, respectively, for the circular arch ASB.

The internal forces of the circular arch ASB are shown in Fig. 3, where *N*, *M*, and *Q* denote mode of axial tension, moment, and shear force, respectively. The relations between these internal forces and displacements are

$$\begin{aligned} N &= \frac{Eh}{R}\left(\frac{dV}{d\theta} + W\right), \\ M &= \frac{Eh^3}{12R^2}\left(\frac{dV}{d\theta} - \frac{d^2W}{d\theta^2}\right), \\ Q &= \frac{Eh^3}{12R^3}\left(\frac{d^2V}{d\theta^2} - \frac{d^3W}{d\theta^3}\right). \end{aligned} \tag{15}$$

Similarly, we have the tangential and normal component of natural mode for the circular arch ASB

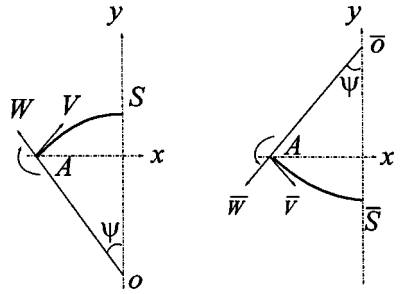


FIG. 4. Displacements and rotation at point A of the arch ASB and the arch $\bar{A}\bar{S}\bar{B}$. *V* and *W* denote, respectively, tangential and normal vibration mode of the arch ASB. $\bar{V}$ and $\bar{W}$ denote, respectively, tangential and normal vibration mode of the arch $\bar{A}\bar{S}\bar{B}$. Displacements and rotation at point A are continuous as the double-circular arch vibrates.

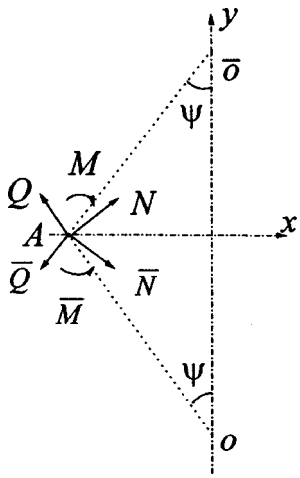


FIG. 5. Internal forces at point A of the arch ASB and the arch $\bar{A}\bar{S}\bar{B}$ are equilibrium in *x*, *y*, and rotation in direction perpendicular to *x* and *y* direction as the double-circular arch vibrates.

$$\bar{V} = \bar{\mathbf{c}}^T\mathbf{G}\mathbf{f}, \tag{16}$$

$$\bar{W} = \bar{\mathbf{c}}^T\mathbf{f}, \tag{17}$$

where $\bar{\mathbf{c}}_{6\times 1}$ is a vector with six unknown constants for arch $\bar{A}\bar{S}\bar{B}$, $\mathbf{f}_{6\times 1}$ is the same solution vector as that in Eq. (9), and $\mathbf{G}_{6\times 6}$ is given in Eq. (13).

Expressions for $\bar{N}$, $\bar{M}$, and $\bar{Q}$, which denote the axial tension, moment, and shear force, respectively, of the mode in the circular arch $\bar{A}\bar{S}\bar{B}$, are the same as (15).

B. Continuity conditions between the circular arch ASB and circular arch $\bar{A}\bar{S}\bar{B}$

Referring to Fig. 4, displacements and rotation angle at point A of arch ASB and arch $\bar{A}\bar{S}\bar{B}$ are continuous. Continuity conditions of displacement are

$$-W\sin\psi + V\cos\psi = -\bar{W}\sin\psi + \bar{V}\cos\psi \quad \text{in } x \text{ direction}, \tag{18}$$

$$W\cos\psi + V\sin\psi = -\bar{W}\cos\psi - \bar{V}\sin\psi \quad \text{in } y \text{ direction}, \tag{19}$$

$$V - \frac{dW}{d\theta} = -\left(\bar{V} - \frac{d\bar{W}}{d\theta}\right) \quad \text{in rotation}, \tag{20}$$

where $\theta = \bar{\theta} = -\psi$. Similarly, continuity conditions of displacements at point B are

$$W\sin\psi + V\cos\psi = \bar{W}\sin\psi + \bar{V}\cos\psi, \tag{21}$$

$$W\cos\psi - V\sin\psi = -\bar{W}\cos\psi + \bar{V}\sin\psi, \tag{22}$$

$$V - \frac{dW}{d\theta} = -\left(\bar{V} - \frac{d\bar{W}}{d\theta}\right), \tag{23}$$

where $\theta = \bar{\theta} = \psi$.

Referring to Fig. 5, internal forces at point A of arch ASB and arch $\bar{A}\bar{S}\bar{B}$ are equilibrium. Continuity conditions of force are

$$N\cos\psi - Q\sin\psi + \bar{N}\cos\psi - \bar{Q}\sin\psi = 0$$

in x direction, (24)

$$N\sin\psi + Q\cos\psi - \bar{N}\sin\psi - \bar{Q}\cos\psi = 0$$

in y direction, (25)

$$M - \bar{M} = 0 \quad \text{in moment,} \tag{26}$$

where $\theta = \bar{\theta} = -\psi$. Similarly, continuity conditions of force at point B are

$$-N\cos\psi - Q\sin\psi - \bar{N}\cos\psi - \bar{Q}\sin\psi = 0, \tag{27}$$

$$N\sin\psi - Q\cos\psi - \bar{N}\sin\psi + \bar{Q}\cos\psi = 0, \tag{28}$$

$$M - \bar{M} = 0, \tag{29}$$

where $\theta = \bar{\theta} = \psi$.

Equations (18)–(29), in total 12 equations, are the continuity conditions at points A and B between arch ASB and arch ASB. Substituting Eqs. (9), (14), (16), and (17) into (18)–(29), and taking (15) into account, we have

$$\mathbf{H}(\omega)\begin{pmatrix}\mathbf{c}\\ \mathbf{c}\end{pmatrix} = \mathbf{0}, \tag{30}$$

where $\mathbf{H}(\omega)_{12\times 12}$ is a matrix.

C. Natural frequencies and natural modes of a double-circular arch

As the double-circular arch vibrates, $(v,w) \neq \mathbf{0}$ in (1) for arch ASB and arch ASB. So, from (2), we have

$$(W, V, \bar{W}, \bar{V}) \neq \mathbf{0}, \tag{31}$$

and considering the expression for W, V, $\bar{W}$, $\bar{V}$ in Eqs. (9), (14), (16), and (17), we have

$$\begin{pmatrix}\mathbf{c}\\ \mathbf{c}\end{pmatrix} \neq \mathbf{0}. \tag{32}$$

Considering (32) and (30), we have

$$\det \mathbf{H}(\omega) = 0. \tag{33}$$

The roots $\omega \in (0, +\infty)$ of (33) are the natural frequencies of the double-circular arch in Fig. 2.

Equation (33) is a transcendental equation in which ω is an unknown quantity. In this paper, these roots in $(0, +\infty)$ are obtained numerically by the bisection method (see Sec. II D). Substituting a root of (33) into (30), (9), (14), (16), and (17), the natural mode corresponding to this root can be found.

D. Calculation of det H(ω)

In the bisection method for searching the roots of (33), it is necessary to calculate value of $\det \mathbf{H}(\omega)$ for a given ω. The process for calculating $\det \mathbf{H}(\omega)$ is as follows: First, for a given ω, we can get the analytical solution of λ^2 in (7). Second, we can get the six roots λ of (7). Third, according to the six roots, we can get analytical expression of solution vector $\mathbf{f}$ in (9) for differential equation (4) or (5). Fourth, we have an expression of general solution (9), (14), for (4), (5), respectively. Fifth, substituting expressions for V and W into (15), we have an expression for the internal forces. For arch ASB, we repeat the process above. Sixth, substituting the analytical expressions for displacements and internal forces into (18)–(29), we have an analytical expression of matrix $\mathbf{H}(\omega)_{12\times 12}$ in (30) for a given ω. Finally, we have value of determinant $\det \mathbf{H}(\omega)$ numerically.

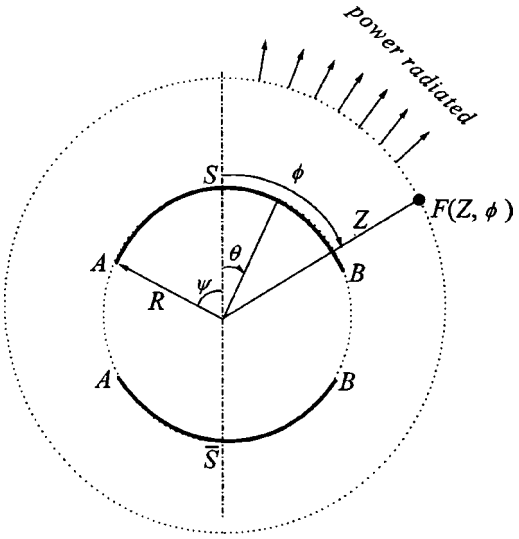


FIG. 6. Model of sound attenuation. As the cylinder with the lens-shaped cross section like a double-circular arch vibrates, its moving surface compresses air periodically to generate a sound wave and radiates acoustic power outward.

III. SOUND ATTENUATION OF A CYLINDER WITH THE LENS-SHAPED CROSS SECTION LIKE A DOUBLE-CIRCULAR ARCH

A. The total power radiated in an unit time

In fact, structure of an ancient Chinese bell is relatively complicated.[1] To analyze its acoustics, we have to do some simplifications. First, since the structure of an ancient Chinese bell in Fig. 1 is approximately a cylinder with a lens-shaped cross section, we consider an infinitely long cylinder with the lens-shaped cross section like a double-circular arch to model the bell's structure. Second, surface vibration of the cylinder has the same natural frequencies and modes as the double-circular arch in Fig. 2. Third, in this treatment, the viscosity of air is neglected. Fourth, we suppose that the sound wave is a cylindrical wave so that the problem is two-dimensional.

As the cylinder vibrates, its moving surface compresses air periodically to generate a sound wave, and it radiates acoustic power outward as shown in Fig. 6. Using polar coordinates (Z, ϕ), the sound field far away from the source is considered where $Z \gg R$.

In Fig. 6, θ denotes coordinate of ASB. According to (2), the radial velocity at point θ of ASB is $LW(\theta)j\omega e^{j\omega t}$, in which L is the displacement amplitude and $W(\theta)$ is the normal component of the natural mode of ASB. The radial velocity at point θ of ASB is $L\bar{W}(\theta)j\omega e^{j\omega t}$, where L is the displacement amplitude and $\bar{W}(\theta)$ is the normal component of the natural mode of ASB. ρ_0 and c denotes the density of air and the velocity of sound in air, respectively, according to

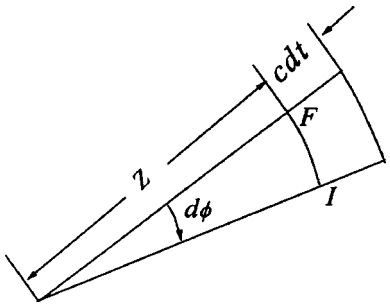


FIG. 7. In an infinitesimal time interval dt, power radiated through surface $F-I$ equals kinetic energy of air particles in square $c\,dt$ by $Z\,d\phi$, where c is sound velocity.

Eq. (7.3.10) on page 363 in Morse.[6] In Fig. 6, the acoustic pressure p and the velocity u of an air particle at $F(Z,\phi)$ far away from the sound source ASB and ASB are

$$p(Z,\phi,t)=\rho_0 c\sqrt{\frac{R}{Z}}e^{j\delta(Z-ct)}\left[\int_{-\psi}^{\psi}f(\phi-\theta)LW(\theta)j\omega\,d\theta+\int_{\pi-\psi}^{\pi+\psi}f(\phi-\theta)L\overline{W}(\pi-\overline{\theta})j\omega\,d\overline{\theta}\right], \tag{34}$$

$$u(Z,\phi,t)=\frac{p}{\rho_0 c}, \tag{35}$$

where $\delta=\omega/c$ and f is shown in Eq. (7.3.8) on page 361 in Morse[6]

$$f(\lambda)=\sqrt{\frac{2}{\pi^3\delta R}}\sum_{n=0}^{\infty}\frac{\cos(n\lambda)}{E_n}\exp\left\{-j\left[\gamma_n+\frac{\pi}{4}(2n+1)\right]\right\}, \tag{36}$$

where E_n and γ_n are determined by

$$J_1(\delta R)=\tfrac{1}{2}E_0\sin\gamma_0,\quad N_1(\delta R)=-\tfrac{1}{2}E_0\cos\gamma_0, \tag{37}$$

$$J_{n+1}(\delta R)-J_{n-1}(\delta R)=2E_n\sin\gamma_n,$$
$$N_{n-1}(\delta R)-N_{n+1}(\delta R)=2E_n\cos\gamma_n\quad(n=1,2,3,\ldots), \tag{38}$$

where $J_n(\,)$ and $N_n(\,)$ are the Bessel and Neumann function of the nth order, respectively.

Substituting (34) into (35), the velocity of an air particle at $F(Z,\phi)$ is

$$u(Z,\phi,t)=j\omega L\sqrt{\frac{R}{Z}}e^{j\delta(Z-ct)}\left[\int_{-\psi}^{\psi}f(\phi-\theta)W(\theta)d\theta+\int_{\pi-\psi}^{\pi+\psi}f(\phi-\theta)\overline{W}(\pi-\overline{\theta})d\overline{\theta}\right]. \tag{39}$$

From Eq. (39), we have the maximum velocity of an air particle at $F(Z,\phi)$

$$u_{\max}(Z,\phi)=\omega L\sqrt{\frac{R}{Z}}\left[\int_{-\psi}^{\psi}f(\phi-\theta)W(\theta)d\theta+\int_{\pi-\psi}^{\pi+\psi}f(\phi-\theta)\overline{W}(\pi-\overline{\theta})d\overline{\theta}\right]. \tag{40}$$

Let's consider the movement of air particles near surface $F-I$ as the sound wave is traveling through $F-I$ in Fig. 7. Before the sound wave reaches $F-I$, air particles near $F-I$ are still. At time t, sound wave reaches $F-I$. After an infinitesimal time interval dt, i.e., at time $t+dt$, sound wave travels through $F-I$ and air particles in volume $Zc\,d\phi\,dt$ get vibrating with maximum velocity $u_{\max}$. The kinetic energy of these air particles equals the power radiated through surface $F-I$ in time interval dt. So, power radiated through surface $F-I$ in an infinitesimal time interval dt is

$$\tfrac{1}{2}\rho_0 Zc\,d\phi\,dt\,u_{\max}^2. \tag{41}$$

Substituting (40) into (41), and integrating (41) with respect to ϕ from 0 to 2π, we have the total power radiated through the large circle with radius Z in Fig. 6 in unit time

$$\frac{d\Pi}{dt}=R\omega^2L^2\Sigma, \tag{42}$$

where

$$\Sigma=\frac{1}{2}\rho_0 c\int_0^{2\pi}\left[\int_{-\psi}^{\psi}f(\phi-\theta)W(\theta)d\theta+\int_{\pi-\psi}^{\pi+\psi}f(\phi-\theta)\overline{W}(\pi-\overline{\theta})d\overline{\theta}\right]^2d\phi, \tag{43}$$

and $d\Pi$ is the acoustic power radiated through the circle with radius Z in time interval dt. Expression (42) is the total power radiated in a unit time from the cylinder with the lens-shaped cross section like a double-circular arch.

B. Attenuation exponential of the vibrating cylinder with the lens-shaped cross section

The kinetic energy of the vibrating cylinder with the lens-shaped cross section like a double-circular arch is

$$T=\tfrac{1}{2}R\omega^2L^2\Omega, \tag{44}$$

where

$$\Omega=\rho h\left[\int_{-\psi}^{\psi}(W^2+V^2)d\theta+\int_{-\psi}^{\psi}(\overline{W}^2+\overline{V}^2)d\overline{\theta}\right], \tag{45}$$

where ρ is density of the cylinder. By neglecting losses of energy by all means except acoustic radiation, we have

$$dT+d\Pi=0, \tag{46}$$

where dT is the decrease of vibration energy of the cylinder in dt. Substituting (42) and (44) into (46), we have

$$\frac{L}{L_0}=e^{-\beta t}, \tag{47}$$

TABLE I. Natural frequency (Hz) of a circular ring. As $\psi=\pi/2$, the double-circular arch becomes a circular ring. Numerical results of frequencies in this paper are compared with analytical results shown in Soedel (Ref. 4) and Timoshenko (Ref. 7).

(Hz)	Analytical solution		Numerical solution
	Soedel (Ref. 4)	Timoshenko (Ref. 7)	This paper
f_1	126.306	126.301	126.27
f_2	357.248	357.229	357.24

TABLE II. Parameters of two kinds of double-circular arches with different total arc length. The total length is sum of arc length ASB and ASB in Fig. 2.

Double-circular arch	Density (kg/m^3) ρ	Young's modulus (GPa) E	Thickness (mm) h	Total length (mm)
No. 1	7.85×10^3	200.6	2.0	$2\times3.14\times100.0$
No. 2			4.0	$2\times3.14\times50.5$

where L and L_0 are the displacement amplitude of the cylinder at present time t and initial time $t=0$, respectively, and where

$$\beta=\frac{\Sigma}{\Omega} \tag{48}$$

is the attenuation exponential of sound.

IV. NUMERICAL RESULTS

A. Natural frequencies for a circular ring

As $\psi=\pi/2$, the double-circular arch in Fig. 2 becomes a circular ring. As Young's modulus $E=200.6\,\mathrm{GPa}$, density $\rho=7.85\times10^3\,\mathrm{kg/m^3}$, thickness $h=2\times10^{-3}\,\mathrm{m}$, and radius $R=1\times10^{-1}\,\mathrm{m}$, the numerical results in this paper on the first and the second natural frequencies are compared with the results in Soedel[4] and Timoshenko[7] in Table I.

B. Two tones of an ancient Chinese bell

Parameters of two kinds of double-circular arches No. 1 and No. 2 in different total arc length are listed in Table II. The first and the second natural modes of a double-circular arch are shown in Fig. 8. Ratios of the second frequency f_2 to the first frequency f_1 of the two kinds of double-circular arches are listed in Table III for different ψ. The numerical results indicate that values of f_2/f_1 for the two kinds of double-circular arches No. 1 and No. 2 remain the same for a given ψ as shown in Table III.

In Fig. 8 and Table III, we can see that

(i) As the central point (SUI position on the bell) is struck, the double-circular arch vibrates at the first frequency f_1. As the side point (GU position on the bell) is struck, the double-circular arch vibrates at the second frequency f_2.

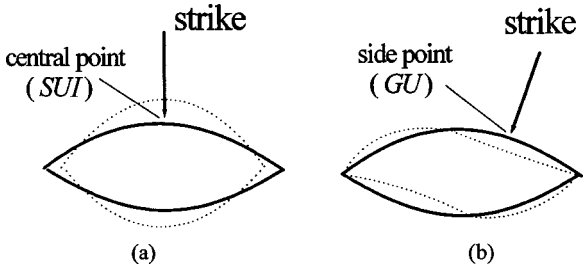


FIG. 8. Two tones of an ancient Chinese bell. As the central point (SUI position on bell) is struck, the double-circular arch vibrates in the first frequency with the first natural mode in (a). As the side point (GU position on bell) is struck, it vibrates in the second frequency with the second natural mode in (b).

(ii) If the cross section is circular, the bell emits simply one pitch as an arbitrary point is struck. If the cross section is lens-shaped, the bell emits two pitches as SUI position and GU position are struck, respectively.

C. Short duration of sound in Chinese bell

For example, we take the cylinder with the lens-shaped cross section like a double-circular arch No. 2 in Table II to simulate the short duration of sound in Chinese bell. Under 20 °C, density of air is $\rho_0=1.293\,\mathrm{kg/m^3}$ and sound velocity in air is $c=340\,\mathrm{m/s}$. Attenuation exponentials β in (47) for the first tone and the second tone are shown in Table IV for different ψ.

In Table IV, we can see that: The smaller the angle ψ is, the bigger the attenuation exponential β is. In other words, the more oblate the bell is, the more quickly the sound echo attenuates. SUI pitch (frequency f_1) and GU pitch (frequency f_2) both have the same property. For a circular cylinder ($\psi=90°$), attenuation exponential β is the minimum. This is why sound echo of an ancient Chinese bell attenuates more quickly than a round bell, which is necessary to musical instruments, just because of its lens-shaped cross section as shown in Fig. 1.

In Table IV, we can see that ψ controls the short duration of sound. According to (47) and Table IV, the sound of the first tone decreases by 10 times (as $L/L_0=0.1$) after it spends approximately 4.9, 3.0, 1.8, and 1.2 s for $\psi=85°$, 75°, 65°, and 55°, respectively, and the sound of the second tone decreases by 10 times (as $L/L_0=0.1$) after it spends approximately 3.9, 1.9, 1.1, and 0.9 s for $\psi=85°$, 75°, 65°, and 55°, respectively.

V. CONCLUSIONS

Although the model proposed by the author is crude, we get some insight into the acoustics of an ancient Chinese

TABLE III. f_2/f_1 for different ψ of two kinds of double-circular arches No. 1 and No. 2. The more oblate a double-circular arch is, the larger the difference between the first and the second frequency.

ψ	90°	85°	80°	75°	70°	65°	60°	55°	50°
f_2/f_1	1.00	1.02	1.05	1.09	1.13	1.18	1.24	1.31	1.39

TABLE IV. Attenuation exponential β (1/s) for different ψ of the cylinder with the lens-shaped cross section like a double-circular arch No. 2 in Table II. The more oblate the cylinder is, the more quickly the sound echo attenuates.

ψ	90°	85°	80°	75°	70°	65°	60°	55°
β for f_1	0.42	0.47	0.60	0.77	0.99	1.27	1.60	1.90
β for f_2	⋯	0.59	0.85	1.20	1.65	2.15	2.57	2.60

bell. The lens-shaped cross section is an important factor for the two-tone bell. By a vibrating double-circular arch, we can explain theoretically two tones of an ancient Chinese bell. By acoustics of a vibrating cylinder with the lens-shaped cross section like a double-circular arch, we can explain the short duration of sound of an ancient Chinese bell. The numerical results of the model in this paper predict that the more oblate the lens-shaped cross section of Chinese bell is, the more quickly the sound attenuates.

However, it is difficult to model the acoustics of an ancient Chinese bell strictly because a three-dimensional problem on vibration and acoustics should be considered. To understand the ancient Chinese bell completely, we should appeal to numerical methods on the vibration and acoustics in three-dimensions.

ACKNOWLEDGMENTS

The author is grateful to Professor J. K. Wu for discussion. The author gratefully acknowledges support from the National Science Foundation of P.R.China through grant number 90205007 and the Science Foundation of Northern Jiaotong University through grant numbers TJ2002J0180 and PD-118.

[1] S. Sinyan, "Acoustics of ancient Chinese bells," Sci. Am. **256**(4), 94–102 (1987).
[2] T. D. Rossing, D. S. Hampton, B. E. Richardson, and H. J. Sathoff, "Vibration modes of Chinese two-tone bells," J. Acoust. Soc. Am. **83**(1), 369–373 (1988).
[3] P. Chidamparam and A. W. Leisse, "Vibration of planar curved beams, rings, and arches," Appl. Mech. Rev. **48**, 467–483 (1993).
[4] W. Soedel, *Vibration of Shells and Plates* (Marcel Dekker, New York, 1981).
[5] P. Chidamparam, "Free vibration and buckling of curved beams subjected to distributed loads," Ph.D. dissertation, Ohio State University (1993).
[6] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), pp. 360–366
[7] S. P. Timoshenko, D. H. Young, and W. Weaver, Jr., *Vibration Problems in Engineering*, 4th ed. (Wiley, New York, 1974).

The broadband social acoustic signaling behavior of spinner and spotted dolphins

Marc O. Lammers[a] and Whitlow W. L. Au
Marine Mammal Research Program, Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

Denise L. Herzing
Florida Atlantic University, Biological Sciences, Boca Raton, Florida 33431

(Received 25 October 2002; revised 23 May 2003; accepted 9 June 2003)

Efforts to study the social acoustic signaling behavior of delphinids have traditionally been restricted to audio-range (<20 kHz) analyses. To explore the occurrence of communication signals at ultrasonic frequencies, broadband recordings of whistles and burst pulses were obtained from two commonly studied species of delphinids, the Hawaiian spinner dolphin (*Stenella longirostris*) and the Atlantic spotted dolphin (*Stenella frontalis*). Signals were quantitatively analyzed to establish their full bandwidth, to identify distinguishing characteristics between each species, and to determine how often they occur beyond the range of human hearing. Fundamental whistle contours were found to extend beyond 20 kHz only rarely among spotted dolphins, but with some regularity in spinner dolphins. Harmonics were present in the majority of whistles and varied considerably in their number, occurrence, and amplitude. Many whistles had harmonics that extended past 50 kHz and some reached as high as 100 kHz. The relative amplitude of harmonics and the high hearing sensitivity of dolphins to equivalent frequencies suggest that harmonics are biologically relevant spectral features. The burst pulses of both species were found to be predominantly ultrasonic, often with little or no energy below 20 kHz. The findings presented reveal that the social signals produced by spinner and spotted dolphins span the full range of their hearing sensitivity, are spectrally quite varied, and in the case of burst pulses are probably produced more frequently than reported by audio-range analyses. [DOI: 10.1121/1.1596173]

PACS numbers: 43.80.Ka [FD]

I. INTRODUCTION

Dolphins (family: Delphinidae) are known to produce and hear sounds well beyond the frequency limits of human hearing. Their use of short, broadband clicks with peak energies between 60 and 120 kHz is well documented in the context of echolocation behavior [see Au (1993) for a review]. In contrast, considerably less has been reported on their use of ultrasonic signaling for communicative purposes, in part because broadband recorders were until recently quite expensive and not very portable (Au *et al.*, 1999). With a few exceptions (Lilly and Miller, 1961; Brownlee, 1983; Dawson, 1991; Rasmussen and Miller, 2002), the vast majority of studies examining dolphin social signals both in captivity and in the field have been restricted in bandwidth to the human-audible range. This has left the full-band characteristics of their signals poorly described and the use of ultrasonic frequency bands for communicative purposes all but unexplored.

Most dolphin species produce two primary types of sounds thought to play a role in social interactions: tonal, frequency-modulated whistles and rapid repetition rate ''burst pulse'' click trains (Herman and Tavolga, 1980). A few species do not produce whistles and are believed to communicate exclusively via pulsed sounds (Dawson, 1991).

How whistles are used in communication is an ongoing topic of debate among researchers (McCowan and Reiss, 2001), but the general consensus is that they play an important role in maintaining contact between dispersed individuals (Janik, 2000a). Burst pulses have also been strongly implicated in communication (Caldwell and Caldwell, 1967; Dawson, 1991; Norris *et al.*, 1994), but their occurrence and functional significance are still only poorly understood. Some authors have suggested they play an important role in agonistic encounters (Caldwell and Caldwell, 1967; Overstrom, 1983; McCowan and Reiss, 1995), while others have proposed they represent ''emotive'' signals in a broader sense (Lilly and Miller, 1961; Herzing, 1988, 1996), possibly representing graded signals (Brownlee, 1983).

Dolphins produce whistles with fundamental frequencies usually in the human audible range (below 20 kHz). These whistles often also have harmonics, which occur at integer multiples of the fundamental and extend beyond the range of human hearing. Harmonics are integral components of tonal signals produced by departures of the waveform from a sinusoidal pattern. To date, no published efforts have been made to explore the occurrence and functional significance of whistle harmonics. It is not presently known how often they occur in dolphin whistles, how far in frequency they extend, how much energy they contain relative to the fundamental, or why harmonics are present in some whistles but not in others.

Little is also known about the properties of burst pulses. The spectral, temporal, and amplitude characteristics of burst

[a]Electronic mail: lammers@hawaii.edu

TABLE I. The characteristics of the fundamental frequency (F_0) of spinner and spotted dolphin whistles. Values are given as mean±standard deviation unless indicated otherwise.

	N	Duration (s)[a]	Mean F_0 frequency (kHz)[a]	FM range of F_0 (kHz)	F_0 max (kHz)[a] mean±S.D. (95% C.I.)	F_0 min (kHz)[a] mean±S.D. (95% C.I.)	% energy in F_0 min, Q_1, *median*, Q_2, max	No. of harmonics
S. longirostris	167	0.66±0.36	13.8±2.3	7.3±3.9	17.4±3.0(16.9–17.9)	10.1±2.5(9.7–10.5)	65.6, 83.3, *93.1*, 95.7, 99.3	1.83±0.87
S. frontalis	220	0.44±0.30	10.9±2.0	7.4±2.9	14.5±2.5(14.1–14.9)	7.1±1.5(6.9–7.3)	47.1, 88.6, *93.5*, 96.9, 99.7	1.99±1.03

[a]$P<0.01$.

pulse click trains have been only marginally explored. Early descriptions in the literature were mostly qualitative, reflecting the subjective aural interpretations and classifications of human listeners (Busnel and Dziedzic, 1966; Caldwell and Caldwell, 1967, 1971). Later efforts were more quantitative, but as with studies of whistles, these also relied primarily on audio-range analyses (Overstrom, 1983; Herzing, 1988, 1996; McCowan and Reiss, 1995; Van Parijs and Corkeron, 2001). Presently, the only quantitative description of burst pulsing at ultrasonic frequencies is for a nonwhistling species, the Hector's dolphin (*Cephalorhynchus hectori*) (Dawson, 1991). Only anecdotal accounts exist about the occurrence of burst pulses above human-audible frequencies for whistling species (Lilly and Miller, 1961; Norris *et al.*, 1994).

In light of the acute hearing sensitivity to ultrasonic frequencies exhibited by all dolphin species tested so far (Nachtigall *et al.*, 2000), the current lack of knowledge about the broadband properties of whistles and burst pulses make it difficult to fully appreciate their design as communicative signals. In this study, we used broadband recording technology adapted for the field to examine the signaling behavior of two commonly studied species of dolphins, the Hawaiian spinner dolphin (*Stenella longirostris*) and the Atlantic spotted dolphin (*Stenella frontalis*). Our objectives here are to (1) report the full-bandwidth properties of the whistles and burst pulses of these two species, (2) present distinctions in their production, and (3) provide an estimate of the bias associated with recording dolphin social signals narrow band (0–20 kHz).

II. METHODOLOGY

A. Data collection

Recordings of dolphin whistles and burst pulses were obtained using a portable broadband data acquisition system (PBDAS), previously described in Au *et al.* (1999). The system employs a laptop computer to operate a National Instruments DAQCard-AI-16E-4 12-bit analog to digital (A/D) converter PCMCIA card. Dolphin signals were digitally sampled at a rate of 260 kHz, providing a Nyquist frequency for all recordings of 130 kHz. Prior to sampling, signals were conditioned by an analog signal processing unit consisting of a high- and low-pass filter and variable gain. The high- and low-pass filters were set at 600 Hz and 100 kHz, respectively, and the gain provided was 60 dB. Recordings were obtained using a custom-built, 2.5-cm-diam spherical hydrophone flat (±3 dB) to approximately 150 kHz with a calibrated sensitivity of −197 dB re 1 μPa.

The A/D card was programmed to sample a single channel continuously into a circular memory buffer in the computer's random access memory (RAM). Incoming signals were detected aurally through headphones and by monitoring a LED bar graph display that revealed the presence and amplitude of signals independent of their frequency. When signals were heard or observed on the LED, the system's operator pressed a trigger switch that initiated the transfer of data from RAM to the computer's hard disk. Usually, about 1 s of pretrigger data and 2 s of posttrigger data were automatically stored with each trigger.

Recordings were obtained from free-ranging spinner dolphins in Hawaii on 18 occasions and Atlantic spotted dolphins in the Bahamas on 17 occasions. Data were collected off the south and west shores of Oahu, Hawaii, during various periods between June 1997 and August 1998 and in the Bahamas along the western edge of the Little Bahama Bank during the month of August in 1999 and 2001. Vessels ranging from 13 to 62 ft in length were used to approach groups of dolphins in each locale. The hydrophone was typically placed 3 m below the surface, between 50 and 200 m in front of moving or milling animals. Recordings were made as groups approached and passed by the vessel. Group sizes ranged from 3 to approximately 100 individuals. The behavioral states of spinner dolphin groups included resting, traveling, and socializing. The same states were observed for spotted dolphins in addition to both daytime and nighttime foraging. Behavioral states were established using the classification method of Norris and Dohl (1980) for spinner dolphins and Herzing (1996) for spotted dolphins.

B. Data analysis

Cool Edit 96™ was used for the initial visualization of recordings. A 1024-point Hanning window was used to plot all sonograms. Subsequent quantitative analyses were carried out using custom-written Matlab™ 5.2 programs. Whistles were analyzed quantitatively if they had harmonics, if their signal to noise (S/N) ratio was sufficient to unambiguously establish their beginning and ending points, and provided they did not overlap in time and frequency with other whistles. The analysis algorithm employed a short-time Fourier transform approach to establish several parameters of interest. These included (1) the properties of the fundamental frequency contour (maximum, minimum, and mean frequencies as well as signal duration); (2) the number of harmonics present and the relative occurrence of each harmonic in the whistle (as a percent of signal duration); (3) the relationship between amplitude modulation of the fundamental (normalized to the maximum amplitude within a signal) and the

presence of a second harmonic; and (4) the amount of energy in each harmonic relative to the fundamental.

Burst pulses were quantitatively analyzed provided the S/N ratio between the peak-to-peak amplitude of clicks and the root mean square (rms) amplitude of the background noise floor was greater than 12 dB. Burst pulses were evaluated with respect to the number of clicks in a train, their interclick interval (ICI), and the peak frequency, center frequency, and rms bandwidth of individual clicks. In addition, the relative energy present above and below 20 kHz was calculated for each click.

Burst pulses were distinguished from echolocation click trains on the basis of their ICIs. Click trains were considered burst pulses if their mean ICI did not exceed 10 ms. This criterion was based on previous work by Lammers *et al.* (2003) and is discussed in more detail later.

III. RESULTS

A. Whistles

A total of 1106 and 557 whistles were collected from spinner and spotted dolphins, respectively. Visual inspection revealed that 61.4% of spinner dolphin whistles and 69.3% of spotted dolphin whistles had one or more higher harmonics. Of the 679 spinner dolphin whistles with harmonics, 167 (24.5%) were deemed appropriate for quantitative analysis. This was the case for 220 (57.0%) of the 386 spotted dolphin whistles with harmonics. Fewer spinner than spotted dolphins whistles met the criteria for quantitative analysis because spinners tended to chorus more (Brownlee, 1983), resulting in multiple overlapping whistles that did not lend themselves well to further analysis.

1. Properties of the fundamental

The contours of the fundamental frequency of spinner and spotted dolphin whistles were roughly equal in their frequency modulation (FM) range (approximately 7 kHz), but spinner dolphin whistles were significantly (two-sample *t*-test; $P<0.01$) longer in duration and higher in their minimum, maximum and mean frequency than those of spotted dolphins (Table I). All fundamentals had most of their energy below 20 kHz. Eight (3.6%) spotted dolphin and 41 (24.5%) spinner dolphin whistles had maximum frequencies of the fundamental that were above 20 kHz. The highest maximum frequency measured for a spotted dolphin whistle fundamental was 27.1 kHz. For spinner dolphins it was 24.9 kHz.

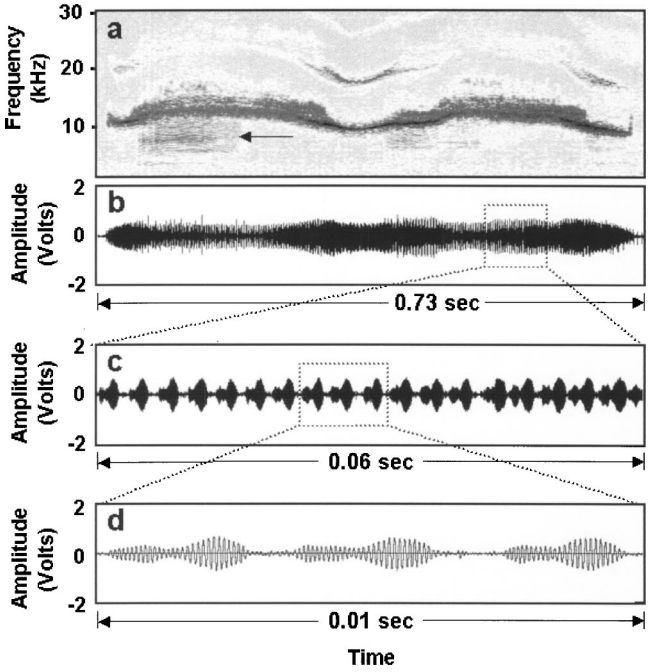


FIG. 1. Amplitude-modulated spotted dolphin whistle showing the sonogram (a), the waveform pattern (b), and a progressive magnification on the pulse-like structure of the signal [(c) and (d)]. The arrow points to a weak concurrent burst pulse.

Spotted dolphins often produced whistles that were qualitatively very distinct from the whistles of spinner dolphins. To human listeners, many spotted dolphin whistles had a coarse or ''raspy'' aural quality that was considerably less pure-tone than those of spinner dolphins. This was because segments of the waveforms of spotted dolphin whistles frequently exhibited such a high degree of amplitude modulation (AM) that they took on pulselike properties (Fig. 1). In the spectral domain, the energy in the fundamental frequency of these kinds of whistles covered a wider band (1–5 kHz) than the more tonal whistles of spinner dolphin (<100 Hz). These rapid AM whistles often occurred either in conjunction with or immediately following burst pulses. Bouts of a specific contour type usually exhibited a very similar AM pattern on consecutive whistles, suggesting that the modula-

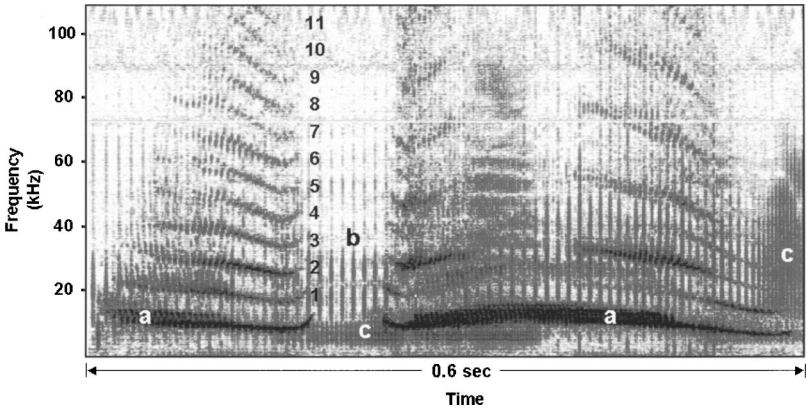


FIG. 2. A sequence of two spotted dolphin whistles (a) mixed with echolocation clicks (b) and burst pulses (c). The first whistle contains 11 harmonics (numbered).

TABLE II. The occurrence, energy content, and amplitude of the second (H_1), third (H_2), fourth (H_3), and fifth (H_4) harmonics relative to the fundamental (F_0). All values except maxima are given as mean±standard deviation.

				% of total signal energy		dB less than F_0	
	Harmonic	*N*	% of signal duration	Mean	Maximum	Mean	Minimum
S. longirostris	H_1	167	74.3±18.9	8.1±5.8	34.1	−11.7±3.7	−2.8
	H_2	96	56.7±27.8	2.2±2.4	13.4	−18.4±5.0	−7.7
	H_3	33	41.0±23.3	0.5±0.5	2.0	−24.6±4.9	−15.6
	H_4	9	37.1±17.5	0.2±0.3	1.0	−28.7±6.6	−19.1
S. frontalis	H_1	220	71.4±25.0	8.4±8.5	43.8	−12.3±5.0	−0.3
	H_2	133	52.1±29.5	2.2±2.6	16.0	−19.0±6.3	−5.2
	H_3	57	47.5±24.6	1.0±0.8	3.1	−20.9±5.5	−13.7
	H_4	29	43.7±22.5	0.5±0.7	3.3	−24.1±5.3	−13.5

tions observed were not simply the byproduct of surface and bottom reflections constructively and destructively interfering with the direct signal [the Lloyd mirror effect (Urick, 1983)]. Approximately 41% of spotted dolphin whistles exhibited this distinct AM pattern on at least part of the signal. Spinner dolphin whistles were also amplitude modulated to a degree, but less than 2% showed the pulsed quality exhibited by spotted dolphins.

2. Occurrence and characteristics of harmonics

The maximum number of harmonics observed was 7 for a spinner dolphin whistle and 11 for that of a spotted dolphin whistle (Fig. 2). More typically, however, the whistles of both species had between one and three harmonics. These did not usually occur throughout the duration of a whistle, but rather varied considerably in the percentage of time they were present in a signal (Table II). In some cases, one or more harmonics occurred throughout the duration of the whistle, while in others harmonics were present during only certain segments of the signal (Fig. 3).

Also variable was the amount of the signal's energy contained in higher harmonics. Although generally about 90% of a whistle's energy was found in the fundamental (F_0), the second harmonic (H_1) could contain up to 34.1% and 43.8% of the total energy of spinner and spotted dolphin signals, respectively. In terms of amplitude, in both species the second harmonic was on average approximately 12 dB lower than the fundamental. However, this difference could be as small as −0.3 dB in some spotted dolphin whistles and −2.8 dB in those of spinner dolphins. The third (H_2), fourth (H_3), and fifth (H_4) harmonics always contained progressively less energy (Table II).

In whistles where one or more harmonics were present during 25%–75% of the signal's duration, there was a significant difference in the relative amplitude of segments that had a second harmonic versus those that did not (paired *t*-test, $P<0.001$). Segments of whistles with harmonics had a median amplitude (normalized to the maximum of each whistle) of 0.65 (S.D.=0.15) for spotted dolphins and 0.67 (S.D.=0.12) for spinner dolphins, while those without were only 0.48 (S.D.=0.19) and 0.53 (S.D.=0.14), respectively. In other words, harmonics consistently occurred at and around the maxima of amplitude modulation in the signal.

B. Burst pulses

Seventy-nine spinner dolphin and 73 spotted dolphin burst pulses were obtained that met the criteria set for quantitative analysis. Burst pulses were produced alone, in bouts or associated with whistles (Fig. 4). Spinner dolphins produced burst pulses that had on average approximately 30 clicks per train with a mean ICI of 3.85 ms (Table III). Spotted dolphin burst pulses averaged about 100 clicks per train with a mean ICI of 3.19 ms. The number of clicks in a burst pulse displayed an approximately bimodal distribution that distinguished burst pulses into low quantity (<70 clicks) and high quantity (>70 clicks) click trains (Fig. 5). Spotted dolphins produced significantly more high quantity burst pulses

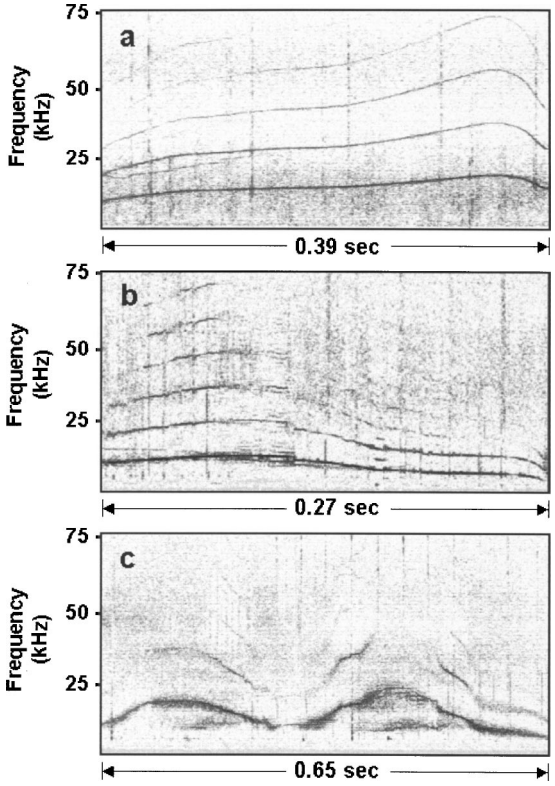


FIG. 3. Examples of the variation found in the harmonic composition of whistles: (a) spinner dolphin whistle with continuous harmonics throughout signal; (b) spotted dolphin whistle with harmonics emphasized on concave portion of the contour; and (c) spotted dolphin whistle with harmonics only on the slopes of the contour.

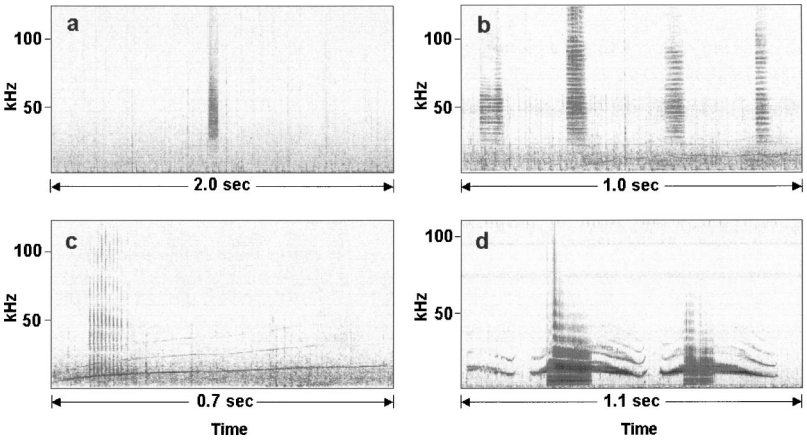


FIG. 4. Examples of the variation in burst pulse occurrence: (a) spinner dolphin burst pulse with minimal energy below 20 kHz occurring alone; (b) sequence of four spinner dolphin burst pulses with minimal energy below 20 kHz occurring within a one second period; (c) simultaneously occurring spinner dolphin whistle and burst pulse with energy both above and below 20 kHz; and (d) simultaneously occurring spotted dolphin whistle and burst pulse with energy mostly below 20 kHz. Individual clicks are not resolved in (a), (b), and (d) due to the inherent time/frequency resolution tradeoff of fast Fourier transforms.

($N=25/73$; $\bar{x}=239$ clicks; max=958 clicks) than spinner dolphins ($N=6/79$; $\bar{x}=116$; max=168 clicks) (chi-square test, $P<0.001$).

Peak and center frequencies for spinner dolphin burst pulses were on average 32.3 and 40.1 kHz, respectively. For spotted dolphins these were somewhat higher at 40.3 and 44.4 kHz, respectively. The rms bandwidths were roughly equivalent at 20.5 kHz for spinner dolphins and 18.1 kHz for spotted dolphins. Spectral energy distribution in clicks did not show any relationship to either the number of clicks in a train or the interclick interval.

Only 17.8% of spinner dolphin and 20.2% of spotted dolphin burst pulses had click trains with peak frequencies below 20 kHz (Table IV). On average, approximately 80% of the total energy in burst pulses was above 20 kHz for both species. Moreover, 39.2% of spinner dolphin and 60.3% of spotted dolphin burst pulses had less than 10% of their total energy below 20 kHz. Thus, while some burst pulses were clearly audible and prominent at sonic frequencies (<20 kHz), most were either barely detectable (aurally or visually on a sonogram) or completely devoid of energy at those frequencies [Figs. 4(a) and (b)]. When we digitally resampled each data file to reflect a 20-kHz bandwidth, we found that 40.5% of spinner dolphin and 30.1% of spotted dolphin burst pulses showed no evidence of being present within that frequency range (Fig. 6).

IV. DISCUSSION

A. Signal characteristics

Our results reveal that the whistles and burst pulses of these two species of dolphin span a broader frequency range than is traditionally reported for delphinids. Although the fundamental frequency contours of their whistles occur mostly in the human-audible range (as is typically assumed), their harmonics routinely reach 50 kHz and beyond. In addition, their burst pulse signals are predominantly ultrasonic, often with little or no energy below 20 kHz.

The spectral measures we have presented must be viewed with caution, however, due to the confounding effects of signal directionality. Dolphin clicks are well known to have directional properties (Au, 1993) and recent findings by Lammers and Au (2003) and Miller (2002) indicate that delphinids also project whistles in a frequency-dependent beam. We could not control for signaler orientation in our study. Therefore, our estimates of the energy content and the occurrence of harmonics and burst pulses are almost certainly underestimates of what these dolphins actually produce on-axis of their transmission beam. On the other hand, the temporal characteristics (number of clicks, interclick interval, etc.) of burst pulses are likely to be more robust to directional variations and can therefore be considered representative of what these two species produce.

1. Whistle harmonics

The majority of whistles recorded had one or more harmonics. Whistles without harmonics were generally fainter than those with them. These may have been signals that were produced by animals oriented away from the hydrophone rather than whistles truly lacking harmonic structure. Additionally, dolphins may also have exerted some control over the harmonic composition of their whistles, modulating their occurrence. Our finding that the presence or absence of a second harmonic was matched to amplitude maxima in the fundamental suggests that harmonics could be spectral byproducts of amplitude modulation. As a dolphin increases signal amplitude it may progressively lose its ability to pro-

TABLE III. The temporal and spectral characteristics of spinner and spotted dolphin burst pulses. All values are given as mean±standard deviation. No inference was attempted to distinguish the broadband spectral measures of the two species due to the potentially confounding effects of signal directionality (see discussion).

	N	No. of clicks/ burst pulse[b]	Mean interclick interval (ms)[a]	Peak frequency (kHz)	Center frequency (kHz)	rms bandwidth (kHz)
S. longirostris	79	29±29	3.85±1.67	32.3±12.5	40.1±12.1	20.5±4.3
S. frontalis	73	103±145	3.19±1.40	40.3±17.8	44.4±16.5	18.1±4.8

[a]$P<0.05$.
[b]$P<0.01$.

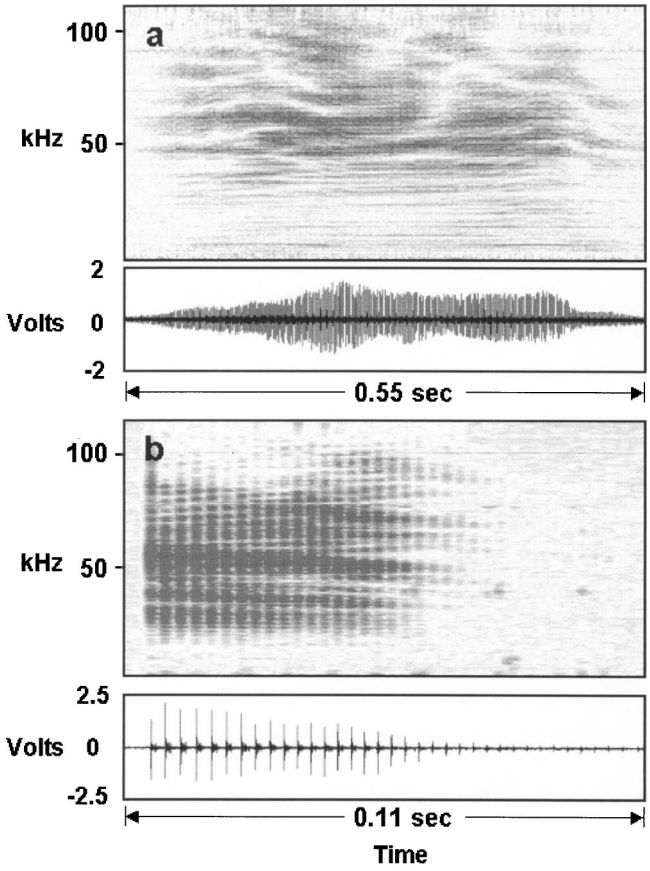


FIG. 5. Examples of high quantity (a) and low quantity (b) spotted dolphin burst pulses. Click train *a* has 255 clicks with mean ICI of 1.7 ms. Click train *b* has 35 clicks with a mean ICI of 2.9 ms.

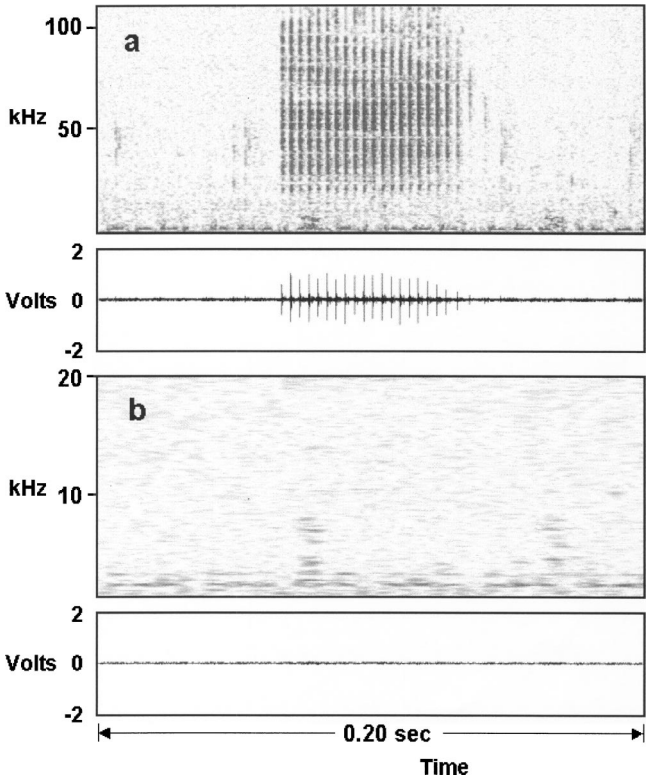


FIG. 6. The same spotted dolphin burst pulse represented both broadband (a) and downsampled to provide a narrow-band perspective (b). Note the lack of any evidence that a signal is present in the narrow-band view.

duce a pure tone, resulting in harmonic distortions. The variation in harmonic composition between and within whistles suggests that directionality and modulation probably both influence their occurrence. Empirical measurements of a whistle beam-pattern with a more detailed examination of how signal amplitude affects harmonic content will provide a more definitive explanation for the variability we observed.

An intriguing result was obtained when we used a delphinid audiogram to infer the way whistles with harmonics might be heard by nearby conspecifics. No audiograms specific to spinner or spotted dolphins presently exist, so we used one recently obtained for a species in the same genus, the striped dolphin (*Stenella coeruleoalba*) (Kastelein *et al.*, 2003). Adjusting the relative amplitude of each harmonic (from Table II) with respect to the striped dolphin's hearing sensitivity at the corresponding frequencies revealed that dolphins should be quite sensitive to harmonics. For almost all the frequencies examined, increased hearing sensitivity more than compensated for the decreased amplitude of harmonics (Table V). In fact, at most harmonic frequencies the levels presumably received by listeners were between 3 and almost 7 dB higher than the fundamental

Whether listening dolphins hear the fundamental or a specific harmonic as the dominant frequency of a whistle depends on the perceived level and the signal-to-noise ratio of each signal component. Given the broad range in relative amplitudes of the harmonics we measured, it is entirely likely that harmonics sometimes are heard as the dominant frequency of a whistle. Figure 7 illustrates such a case. A whistle is presented as it was first recorded using a hydrophone with a flat frequency response and then bandpass filtered to approximate the striped dolphin's hearing sensitivity. In the filtered whistle the second and third harmonics are in fact higher in amplitude than the fundamental. We can conclude from this that how listening animals actually hear a whistle is greatly influenced by the signal's full spectral makeup, which is in turn determined by the listener's relative position to the emitting beam and perhaps also by active modulation on the part of the signaler.

TABLE IV. Narrow-band (20 kHz) versus broadband (130 kHz) bias of burst pulse spectral content and energy. Values presented as number of signals and % out of *N* or as mean±standard deviation.

	N	No. w/peak frequency below 20 kHz	% energy above 20 kHz	No. w/less than 10% energy below 20 kHz	No. not detectable narrow-band (<20 kHz)
S. longirostris	79	16 (20.2%)	78.7±19.6	31 (39.2%)	32 (40.5%)
S. frontalis	73	13 (17.8%)	80.4±27.8	44 (60.3%)	22 (30.1%)

TABLE V. Comparison of the relative amplitude of harmonics measured from the data (assuming equal sensitivity across all frequencies) and corrected for the hearing sensitivity of *S. coeruleoalba*.

			Mean amplitude of harmonics relative to F_0 (in dB)	
		Mean frequency (kHz)	Equal sensitivity across frequencies (from Table II)	Adjusted with *S. coeruleoalba* audiogram
S. longirostris	F_0	13.8	···	···
	H_1	27.6	−11.7	+3.3
	H_2	41.4	−18.4	+6.6
	H_3	55.2	−24.6	+2.4
	H_4	69.0	−28.7	−3.7
S. frontalis	F_0			
	H_1	10.9	−12.3	−0.3
	H_2	21.8	−19.0	+4.0
	H_3	32.7	−20.9	+6.1
	H_4	43.6	−24.1	+4.9

Whether whistle harmonics play a role in communication among dolphins is a question open for debate. To human listeners, harmonics provide the perceptual quality of timbre and offer qualitative cues that serve to distinguish between otherwise very similar sound patterns (e.g., the same word spoken by different individuals) (Handel, 1989). In many taxa, including other primates, harmonics are often associated with individual identity cues (Masters, 1991; Caudron *et al.*, 1998; Jouventin *et al.*, 1999; Charrier *et al.*, 2002). However, in dolphin signals the directionality of high-frequency spectral features probably result in whistle harmonics being poor indicators of identity due to azimuth-dependent degradation. On the other hand, cueing by listeners on the orientation-dependent amplitude of harmonics could play an important role in promoting group cohesion by revealing the orientation of signalers, as suggested by Miller (2002) and Lammers and Au (2003). Discrimination experiments with captive animals are needed to test whether dolphins in fact attend to changes in the harmonic structure of whistles.

2. Burst pulses

a. Energy distribution. The finding that burst pulses have the majority of their energy at ultrasonic frequencies is novel but comes as little surprise, given the properties of echolocation clicks (see Au, 1993). The burst pulse waveforms we recorded could not be readily distinguished from the other, presumably echolocation clicks that were obtained. Relatively few signals matched the traditional audio-range descriptions given for burst pulses, which often report peak frequencies of a few hundred to a few thousand hertz (e.g., Caldwell and Caldwell, 1967; McCowan and Reiss, 1995; Herzing, 1996). Only a minority of signals had peak frequencies below 20 kHz. However, numerous burst pulses displayed ''breaks'' in their energy distribution, reflecting a rippled spectrum caused by closely spaced surface and bottom reflections (Au, 1993). These often had a small energy peak in the audible range with considerably more energy distributed well above 20 kHz (Fig. 8). We presume that many of the previous narrow-band descriptions of burst

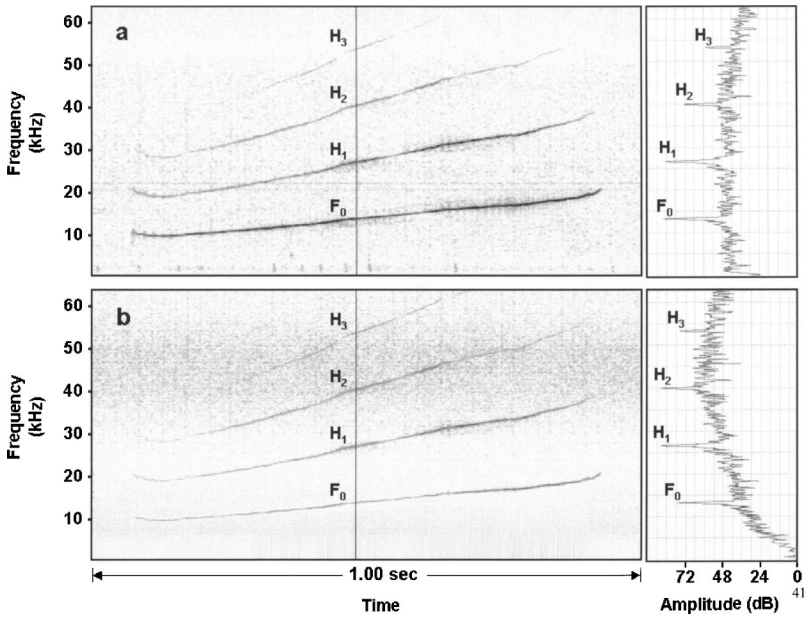


FIG. 7. Sonogram and power spectrum of a spinner dolphin whistle as it was recorded with a hydrophone flat in frequency response (a), and subsequently filtered with a frequency sensitivity curve approximating a striped dolphin's audiogram (b) (after Kastelein *et al.*, 2003). The solid vertical line represents where the 1024-point power spectrum was measured. Represented on the sonogram and power spectrum are the fundamental frequency (F_0), as well as the second (H_1), third (H_2), and fourth (H_3) harmonics. Note that the fundamental is approximately 10 dB lower that the second and third harmonics in the power spectrum of the filtered signal.

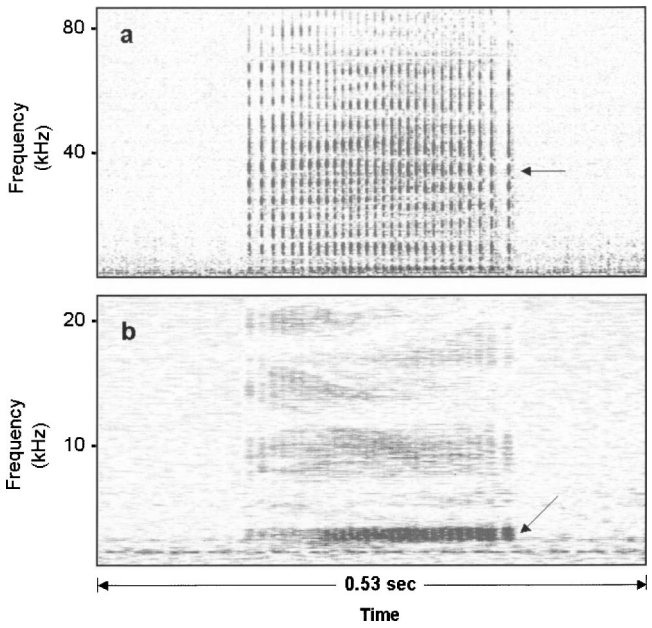


FIG. 8. A spotted dolphin burst pulse represented both broadband (a) and narrow-band (b). The banded pattern of individual clicks represents a rippled spectrum with multiple energy peaks of varying amplitude. The arrows indicate the location of the peak frequency in each case.

pulses may have been accounts of only these lower spectral peaks. In all likelihood, if a broader spectral analysis had been employed in earlier studies, considerably more energy would have been noted at the higher frequency bands.

Of particular interest is the proportion of burst pulses that had little or no energy at sonic frequencies. The fact that 30%–40% of burst pulses were undetectable below 20 kHz implies that they are probably a more common form of social signaling among delphinids than previously suspected. In our recordings, few whistling periods were observed that did not also coincide with at least some burst pulsing (sonic or ultrasonic). Consequently, it is our conclusion that at least in these two species burst pulses probably play an equally important if perhaps different social signaling role as do whistles.

b. Social signaling or echolocation? Given the similarities that exist between burst pulses and other click trains, a perplexing question invariably arises: what is the distinction between click trains presumably used for communication and those used for echolocation? The existing literature on this topic is vague because burst pulses have been traditionally discussed in terms of their sonic properties (Herman and Tavolga, 1980; Popper, 1980; Overstrom, 1983; Herzing, 1996; Van Parijs and Corkeron, 2001). In our analysis, we considered a signal to be a burst pulse if the interclick intervals did not exceed 10 ms. This criterion was based on a previous effort that examined patterns of click production in spinner dolphins which found a bimodal distribution of mean interclick intervals separated at around 10 ms (Lammers *et al.*, 2003). This bimodal distribution was interpreted as being indicative of two distinct patterns of click train production: a burst pulse pattern with interclick intervals consistently less than 10 ms and a sonar click train pattern with intervals greater than approximately 15 ms.

Admittedly, this distinction is not necessarily definitive because click trains are sometimes produced that begin with long interclick intervals (10–100+ ms) and end with very short ones (1.5–9 ms). These are often observed when animals are foraging and presumably echolocating on prey (Herzing, 1996). However, a rationale can be presented for the 10-ms distinction, given that it is not clear whether dolphins process the returning echoes from click trains with very short intervals. Experiments with free-swimming bottlenose dolphins (*Tursiops truncatus*) have shown that an echo-processing lag time of between 15 and 45 ms is always associated with successive clicks produced by animals echolocating on targets further than 0.4 m away (Evans and Powell, 1967; Au, 1993). As animals close in on a target ($<$0.4 m), interclick intervals as low as 2.5 ms have been observed (Evans and Powell, 1967; Morozov *et al.*, 1972), but whether dolphins are processing more than one echo at a time, selecting specific echoes or even using the returning echoes at all, is not known.

More telling perhaps is the fact that, even at close range, there is always a gradual progression towards shorter click intervals as the dolphin approaches a target (Evans and Powell, 1967; Morozov *et al.*, 1972). Therefore, trains characterized by variable ICIs considerably greater and less than 10–15 ms probably represent a type of echolocation not yet well understood rather than a functionally separate class of signals. However, no evidence presently exists to indicate that click trains that begin, persist, and end with interclick intervals less than 10 ms are ever used in contexts identified as echolocation. This does not exclude the possibility that an echolocation function may be associated, but rather implies that burst pulses, as defined here, form a class of click trains distinct in occurrence from those typically linked to echolocation. Their frequent incidence in highly social contexts (such as agonistic encounters) considered with their absence in controlled echolocation contexts strongly suggests that their primary function is likely communicative. In other words, burst pulses are probably signals intended to be heard by nearby animals. Additional broadband recordings coupled with underwater visual observations of socializing and echolocating animals will be needed to further explore the distinction between burst pulse and echolocation click train production, as well as to establish how burst pulses function in communication.

c. Nomenclature. Given that burst pulses are predominantly ultrasonic signals, we propose that the practice of discussing them in terms of their aural qualities should to be reconsidered in favor of a more quantitative approach. Terms such as ''squawks,'' ''squeaks,'' ''creaks,'' and ''yelps'' commonly used to describe and distinguish burst pulses can result in misleading conclusions, as they primarily describe the subjective impressions of time separation pitch experienced by human listeners (McCellan and Small, 1965). Dolphins have better auditory time resolution capabilities than humans (Vel'min and Dubrovskiy, 1976; Ketten, 1992), so attempting to classify burst pulses on the basis of human-perceived aural qualities likely misrepresents how dolphins hear the sounds. As a start, forming classes on the basis of temporal characteristics, such as interclick intervals and total number of

clicks, would lead to more meaningful comparative discussions between authors.

B. Species distinctions

Although some differences were noted in the broadband spectral properties of spinner and spotted dolphin signals, no attempt was made to formally define these distinctions due to the potentially confounding influences of signal directionality mentioned previously. However, differences in the more narrow-band and temporal properties of their whistles and burst pulses merit further comment.

1. Frequency of the fundamental

The fundamental whistle contours of spinner dolphins were on average 3 kHz higher in mean and maximum frequency than those of spotted dolphins. This can in part be explained by the disparity of their adult sizes. Hawaiian spinner dolphins range in length between 170 and 200 cm (Norris *et al.*, 1994), whereas Atlantic spotted dolphins are between 166 and 229 cm in length (Perrin *et al.*, 1994). Ding *et al.* (1995) established that a linear relationship exists between the typical body length of delphinid species and the maximum frequency of their whistle fundamentals. According to the formula they derived, the maxima of spinner dolphin whistles should fall between 14.9 and 15.9 kHz and those of spotted dolphins between 14.0 and 16.1 kHz. The fact that spinner dolphins consistently produced whistles higher in frequency than expected (95% C.I.=16.9–17.9 kHz) whereas spotted dolphins conformed to predicted values (95% C.I.=14.1–14.9 kHz) may reflect a response on the part of spinners to the noisy inshore waters typically found in Hawaii. Most of our recordings of spinner dolphins were made in shallow waters (<100 ft) near reefs where snapping shrimp (family: Alpheidae) produced a continuous stream of broadband clicks (Au and Banks, 1998). The resulting background noise in these areas was greatest below 10 kHz and decayed with increasing frequency. Bahamian waters were comparatively quieter, as spotted dolphins primarily occupied vast stretches of sandy bottom where snapping shrimp were rare. By producing whistles at higher frequencies, spinner dolphins improve their S/N ratio and may thus increase the range or ''active space'' of their signals (Janik, 2000b).

2. Amplitude modulation

Interference resulting from surface and bottom reflections can confound differences in the AM properties of whistles. However, the clear distinctions we found between the two species were sufficiently consistent to make us believe they were not simply artifacts of sound propagation. The pulselike qualities that characterized many spotted dolphin whistles support a contention previously made by Murray *et al.* (1998) that delphinid sounds are produced as a continuum of signals graded between exponentially damped sinusoidal pulses (clicks) and continuous sinusoidal tones (whistles). The AM whistles of spotted dolphins appear to be intermediary signals between the two ends of this continuum. To our knowledge, this is the first description of such signals obtained from the wild.

Whistles characterized by tonal pulses were produced commonly by spotted dolphins but rarely by spinners. This says something about the type of information that might be associated with the whistles of each species. Amplitude modulations in a tone degrade and become indistinct with increasing distance from the source (Urick, 1983). Spotted dolphin whistles with a high degree of AM may therefore be used more for communication with nearby listeners than for long-range signaling between individuals. This is consistent with their social structure, as spotted dolphins in the Bahamas occur in fairly small groups of 2–15 animals characterized by strong relationships between individuals (Herzing and Brunick, 1997). Communicating via AM cues about a behavioral, emotive, or referential condition to a nearby pairmate or kin could be a primary function of their whistles. In contrast, spinner dolphins typically occur in much larger, dynamic aggregations of 20 to 100+ individuals (Östman, 1994; Lammers, 2003) and rely heavily on the group's cohesion and coordination for defense against predators and cooperative foraging at night (Norris *et al.*, 1994; Benoit-Bird and Au, 2001). Spinner dolphin groups often travel with individuals spread over many hundreds of meters. Based on their social structure, communicating over longer ranges using signals designed to favor localization of the signaler is probably more adaptive and thus results in whistles produced with less AM but longer duration.

3. Click number and ICI

The typical number of clicks and the average interclick interval in burst pulses differed between the two species. Both produced relatively short burst pulses with less than 70 clicks, but spotted dolphins also produced many high quantity (>70 clicks) burst pulses whereas spinner dolphins did not. In addition, the interclick intervals of spotted dolphin burst pulses were somewhat shorter than those of spinner dolphin burst pulses.

Our surface behavioral observations were too limited in detail to match signals with the occurrence of specific behaviors; therefore our ability to speculate on the significance of these differences is limited. We suspect, however, that they may be tied to the daily patterns of activity characteristic of each species. Spotted dolphins were generally active during the day, spending much of their time foraging and socializing. Spinner dolphins, on the other hand, predominantly engaged in these activities in the late afternoon and at night, choosing to rest quietly during the middle of the day (Norris *et al.*, 1994). It is possible that spinner dolphins produced more high quantity burst pulses at night. However, since all spinner dolphin recordings were obtained during the day, we presently cannot address this question further.

Similarly, the difference in interclick intervals may also be a function of behavioral state. If burst pulses are indicators of emotive disposition, as has been suggested, then it is to be expected that variations would be tied to behavioral activities. Attempting to classify the fine scale temporal characteristics of burst pulses on the basis of their occurrence in specific behavioral contexts could be a fruitful avenue of future investigation.

V. CONCLUSION

If we presume that dolphins pay attention to the harmonic composition of whistles and if we accept that burst pulses play an important social role, then the evidence presented here indicates that there is considerably more to the social acoustic signaling behavior of spinner and spotted dolphins than meets the human ear. In future efforts to better understand the function of whistles and burst pulses we will need to more fully explore and appreciate their design. This will require that we take into account their broadband patterns of production and adopt methodologies that reflect the dolphin's auditory acuity. Recording signals in a manner consistent with how they are produced and ultimately heard by their intended listeners will be an important key to future efforts to accurately match them with their social context. Doing this will be as important for efforts to study spinner and spotted dolphin signaling as for those involving other species of delphinids.

ACKNOWLEDGMENTS

We thank Roland Aubauer for his help with the development of the PBDAS, Mandy Hapner and Danielle Lanyard for their assistance with data collection in Hawaii, and Kelly Moewe and the crew and staff of the Wild Dolphin Project for their help in the Bahamas. Paul Nachtigall provided important support for this work as well as thoughtful inputs on earlier drafts of this manuscript. Lenny Freed, George Losey, Joe Mobley, and Julie Oswald also provided valuable commentary. This work was conducted under NMFS Permit No. 1000-1617 and under the authority of the Bahamian Ministry of Fisheries. Funding was provided in part through a grant from the Shedd Aquarium as well as a grant from NOAA project #R/FM-7, sponsored by the University of Hawaii Sea Grant College Program under Institutional Grant No. NA86RG0041. This is HIMB publication #1158.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Au, W. W. L., and Banks, K. (**1998**). ''The acoustics of the snapping shrimp *Synalpheus parneomeris* in Kaneohe Bay,'' J. Acoust. Soc. Am. **103**, 41–47.

Au, W. W. L., Lammers, M. O., and Aubauer, R. (**1999**). ''A Portable Broadband Data Acquisition System for field studies in bioacoustics,'' Marine Mammal Sci. **15**, 526–531.

Benoit-Bird, K., and Au, W. W. L. (**2001**). ''Foraging behavior of the Hawaiian spinner dolphin, *Stenella longirostris*,'' Abstracts of the 14th Biennial Conference on the Biology of Marine Mammals, Vancouver, B. C., December 2001.

Brownlee, S. M. (**1983**). ''Correlations between sounds and behavior in wild Hawaiian spinner dolphins (*Stenella longirostris*),'' Masters thesis, University of California Santa Cruz.

Busnel, R. G., and Dziedzic, A. (**1966**). ''Acoustic signals of the Pilot whale *Globicephala melaena*, *Delpinus delphis* and *Phocoena phocoena*,'' in *Whales, Dolphins and Porpoises*, edited by K. S. Norris (Univ. of California, Berkley, CA), pp. 607–648.

Caldwell, M. C., and Caldwell, D. K. (**1967**). ''Intraspecific transfer of information via the pulsed sound in captive Odontocete Cetaceans,'' in *Animal Sonar Systems: Biology and Bionics*, edited by R. G. Busnel (Laboratoire de Physiologie Acoustic, Jouy-en-Josas, France), pp. 879–936.

Caldwell, M. C., and Caldwell, D. K. (**1971**). ''Underwater pulsed sounds produced by captive spotted dolphins, *Stenella plagiodon*,'' Cetology **1**, 1–7.

Caudron, A. K., Kondakov, A. A., and Siryanov, S. V. (**1998**). ''Acoustic structure and individual variation of grey seal (*Halichoerus grypus*) pup calls,'' J. Mar. Biol. Assoc. U.K. **78**, 651–658.

Charrier, I., Mathevon, N., and Jouventin, P. (**2002**). ''How does a fur seal mother recognize the voice of her pup? An experimental study of *Arctocephalus tropicalis*,'' J. Exp. Biol. **205**, 603–612.

Dawson, S. M. (**1991**). ''Clicks and communication: the behavioural and social contexts of Hector's dolphin vocalizations,'' Ethology **88**, 265–276.

Ding, W., Wursig, B., and Evans, W. E. (**1995**). ''Comparisons of whistles among seven odontocete species,'' in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein and J. A. Thomas (De Spil, Woerden, Netherlands), pp. 299–323.

Evans, W. E., and Powell, B. A. (**1967**). ''Discrimination of different metallic plates by an echolocating delphinid,'' in *Animal Sonar Systems: Biology and Bionics*, edited by R. G. Busnel (Laboratoire de Physiologie Acoustic, Jouy-en-Josas, France), pp. 363–382.

Handel, S. (**1989**). *Listening: An Introduction to the Perception of Auditory Events* (MIT, Cambridge, MA).

Herman, L. M., and Tavolga, W. N. (**1980**). ''The communications systems of cetaceans,'' in *Cetacean Behavior: Mechanisms and Function*, edited by L. M. Herman (Wiley–Interscience, New York), pp. 149–209.

Herzing, D. L. (**1988**). ''A Quantitative description and behavioral association of a burst-pulsed sound, the squawk, in captive bottlenose dolphins, *Tursiops truncatus*,'' Masters thesis, San Francisco State University.

Herzing, D. L. (**1996**). ''Vocalizations and associated underwater behavior of free-ranging Atlantic spotted dolphins, *Stenella frontalis* and bottlenose dolphin, *Tursiops truncatus*,'' Aqua. Mamm. **22**, 61–79.

Herzing, D. L., and Brunnick, B. J. (**1997**). ''Coefficients of association of reproductively active female Atlantic spotted dolphins, *Stenella frontalis*,'' Aqua. Mamm. **235**, 155–162.

Janik, V. M. (**2000a**). ''Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*),'' Science **289**, 1355–1357.

Janik, V. M. (**2000b**). ''Source levels and the estimated active space of bottlenose dolphin (*Tursiops truncatus*) whistles in the Moray Firth, Scotland,'' J. Comp. Psychol. **186**, 673–680.

Jouventin, P., Aubin, T., and Lengagne, T. (**1999**). ''Finding a parent in a kingpin penguin colony: The acoustic system of individual recognition,'' Anim. Behav. **57**, 1157–1183.

Kastelein, R. A., Hagedoorn, M., Au, W. W. L., and deHaan, D. (**2003**). ''Underwater audiogram of a striped dolphin (*Stenella coeruleoalba*) measured with narrow-band frequency-modulated signals,'' J. Acoust. Soc. Am. **113**, 1130–1137.

Ketten, D. R. (**1992**). ''The marine mammal ear: Specializations for aquatic audition and echolocation,'' in *The Evolutionary Biology of Hearing*, edited by D. Webster, R. Fay, and A. N. Popper (Springer-Verlag, New York), pp. 717–754.

Lilly, J. C., and Miller, A. M. (**1961**). ''Sounds emitted by the bottlenose dolphin,'' Science **133**, 1689–1693.

Lammers, M. O. (**2003**). ''The behavior and broadband acoustic signaling of Hawai'ian spinner dolphins (*Stenella longirostris*),'' Ph.D. dissertation, University of Hawai'i.

Lammers, M. O., and Au, W. W. L. (**2003**). ''Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*): A signal feature to cue direction of movement?'' Marine Mammal Sci. **19**, 249–264.

Lammers, M. O., Au, W. W. L., Aubauer, R., and Nachtigall, P. E. (**2003**). ''A comparative analysis of echolocation and burst-pulse click trains in *Stenella longirostris*,'' in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (University of Chicago, Chicago, IL), pp. 414–419.

Masters, J. C. (**1991**). ''Loud calls of *Galago crassicaudatus* and *Galago garnettii* and their relation to habitat structure,'' Primates **32**, 153–168.

McCellan, M. E., and Small, A. M. (**1965**). ''Time separation pitch associated with correlated noise bursts,'' J. Acoust. Soc. Am. **38**, 142–143.

McCowan, B., and Reiss, D. (**1995**). ''Maternal aggressive contact vocalizations in captive bottlenose dolphins (*Tursiops truncatus*): wide-band, low frequency signals during mother/aunt-infant interactions,'' Zoo Biol. **14**, 293–309.

McCowan, B., and Reiss, D. (**2001**). ''The fallacy of 'signature whistles' in bottlenose dolphins: a comparative perspective of 'signature information' in animal vocalizations,'' Anim. Behav. **62**, 1151–1162.

Miller, P. J. O. (**2002**). ''Mixed-directionality of killer whale stereotyped calls: a direction-of-movement cue?'' Behav. Ecol. Sociobiol. **52**, 262–270.

Morozov, B. P., Akapiam, A. E., Burdin, V. I., Zaitseva, K. A., and Solovykh, Y. A. (**1972**). ''Tracking frequency of the location signals of dolphins as a function of distance to the target,'' Biofizika **17**, 139–145.

Murray, S. O., Mercado, E., and Roitblat, H. L. (**1998**). ''Characterizing the

graded structure of false killer whale (*Pseudorca crassidens*) vocalizations,'' J. Acoust. Soc. Am. **104**, 1679–1688.
Nachtigall, P. E., Lemonds, D. W., and Roitblat, H. L. (**2000**). ''Psychoacoustic studies of dolphin and whale hearing,'' in *Hearing By Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 330–363.
Norris, K. S., and Dohl, T. P. (**1980**). ''The structure and function of cetacean schools,'' in *Cetacean Behavior: Mechanisms and Function*, edited by L. M. Herman (Wiley–Interscience, New York), pp. 211–261.
Norris, K. S., Würsig, B., Wells, R. S., and Würsig, M. (**1994**). *The Hawaiian Spinner Dolphin* (Univ. of California Press, Berkeley, CA).
Overstrom, N. A. (**1983**). ''Association between burst-pulse sounds and aggressive behavior in captive Atlantic bottlenose dolphins (*Tursiops truncatus*),'' Zoo Biol. **2**, 93–103.
Östman, J. S. O. (**1994**). ''Social organization and social behavior of Hawaiian spinner dolphins (*Stenella longirostris*),'' Ph.D. dissertation, University of California, Santa Cruz.
Perrin, W. F., Caldwell, D. K., and Caldwell, M. C. (**1994**). ''Atlantic spotted dolphin, *Stenella frontalis* (G. Cuvier, 1892),'' in *Handbook of Marine Mammals. Volume 5: The First Book of Dolphins*, edited by S. H. Ridgeway and R. Harrison (Academic, London), pp. 173–190.
Popper, A. N. (**1980**). ''Sound emission and detection by delphinids,'' in *Cetacean Behavior: Mechanisms and Functions*, edited by L. M. Herman (Wiley-Interscience, New York), pp. 1–52.
Rasmussen, M. H., and Miller, L. A. (**2002**). ''Whistles and clicks from white-beaked dolphins, *Lagenorhynchus albirostris*, recorded in Faxaflói Bay, Iceland,'' Aqua. Mamm. **28**, 78–89.
Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw–Hill, New York).
Van Parijs, S. M., and Corkeron, P. J. (**2001**). ''Vocalization and behavior of Pacific humpback dolphins, *Sousa chinensis*,'' Ethology **107**, 701–716.
Vel'min, V. A., and Dubrovskiy, N. A. (**1976**). ''The critical interval of active hearing in dolphins,'' Sov. Phys. Acoust. **2**, 351–352.

Acoustic detection of manatee vocalizations

Christopher Niezrecki,[a] Richard Phillips, and Michael Meyer
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611-6250

Diedrich O. Beusse
College of Veterinary Medicine, University of Florida, PO Box 100126, Gainesville, Florida 32610-0126

(Received 20 December 2002; revised 6 June 2003; accepted 16 June 2003)

The West Indian manatee (*trichechus manatus latirostris*) has become endangered partly because of a growing number of collisions with boats. A system to warn boaters of the presence of manatees, that can signal to boaters that manatees are present in the immediate vicinity, could potentially reduce these boat collisions. In order to identify the presence of manatees, acoustic methods are employed. Within this paper, three different detection algorithms are used to detect the calls of the West Indian manatee. The detection systems are tested in the laboratory using simulated manatee vocalizations from an audio compact disk. The detection method that provides the best overall performance is able to correctly identify ~96% of the manatee vocalizations. However, the system also results in a false alarm rate of ~16%. The results of this work may ultimately lead to the development of a manatee warning system that can warn boaters of the presence of manatees. [DOI: 10.1121/1.1598196]

PACS numbers: 43.80.Ka, 43.60.Bf [WA]

I. INTRODUCTION

In recent years the West Indian manatee (*trichechus manatus latirostris*) has become endangered because of a growing number of fatal collisions with boats. Manatees are routinely identified by their unique scar patterns caused by boat strikes. The boat strikes have resulted in a significant number of manatee deaths and injuries, as well as an increased number of boater ''idle-speed'' or ''no-wake'' zones within the waters of the southeastern United States (see Fig. 1). It is of great interest to the general public to eliminate these manatee impacts as well as easing boating restrictions. Likewise, it is also important to minimize the number of collisions between manatees and boats. Unfortunately, these two constraints are typically in opposition to each other. One possible solution that has recently been proposed by the authors is to develop a cost-effective method for signaling to boaters that manatees are present in the immediate vicinity of heavily used waterways. Thus, the boaters should slow to idle speed and maintain a sharp lookout until clear of the affected area. As a first step in the chain toward achieving this long-term goal, this work focuses on the detection of manatees by means of acoustic vocalization.

Schevill and Watkins were two of the first researchers to describe the vocalizations of the Florida manatee (Schevill and Watkins, 1965). Likewise, Hartman reported their vocabulary to be highly variable and to include several categories of vocalizations (squeal, squeaks, and screams) that were produced in a variety of unrelated circumstances (Hartman, 1969). Steel performed a detailed categorization of manatee vocalizations. Nine different categories of adult vocalizations were established. Additionally, differences in manatee calls generally correlated to differences in animal age, gender, and size (Steel, 1982). The most extensive library of manatee recordings to date was created by O'Shea between 1981 and 1984 working with the United States Geological Survey (O'Shea, 1981–1984). These recordings are used to develop and quantify the performance of the detection algorithms presented in this work. Other researchers studied the rate at which manatee calls are made. Apart from when they are feeding, manatees vocalize approximately 1 to 5 times within a 5-min period (Bengston and Fitzgerald, 1985; Nowacek *et al.*, 2003). Within a protected channel they are not likely to be feeding. Therefore, as long as a manatee warning device remains active for a few minutes, this rate of vocalization should be adequate for detection and location of manatees. Richard-Clark compared the vocalization differences between the breeding population of manatees on the west and east coast of Florida. The results indicate that manatees on the east coast vocalize at slightly higher frequencies than those on the west coast, implying that there may be a barrier between east and west coast breeding populations (Richard-Clark, 1991). Other researchers studied the hearing of manatees and created the first manatee audiogram. The research suggests that manatees' hearing is the most sensitive in a frequency range between 16 and 18 kHz (Gerstein *et al.*, 1999). The research led to the development of a possible manatee warning device that can be used to warn manatees that a boat is approaching (Gerstein, 2002). Other researchers also investigated the calls of the Amazon manatee, *Trichechus inunguis*. The vocalizations were found to be very similar in most respects except that they differ in duration (Evans and Herald, 1970; Sonoda and Takemura, 1973). The vocalization characteristics among different sexes and age classes of Amazon manatees were also investigated (Sousa-Lima *et al.*, 2002). Additionally, several researchers have studied the vocalizations of the dugong. The vocalization of the dugong was also found to be similar to the manatee (Nair and

[a]Electronic mail: niezreck@ufl.edu

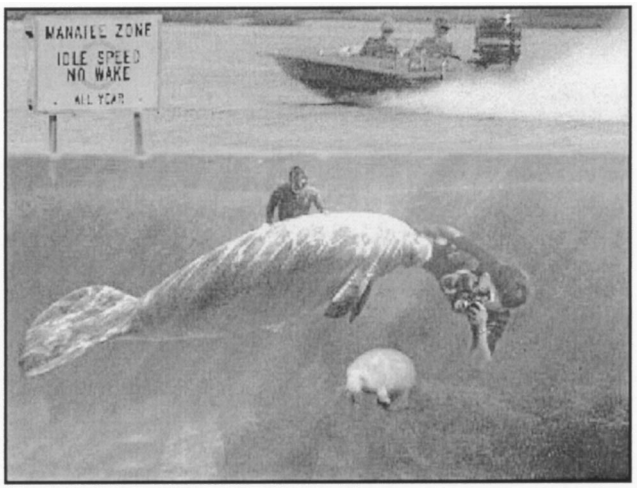


FIG. 1. Manatee warning sign with boater (http://www.xtalwind.net/~cfa/, 2002).

Lal Mohan, 1975; Anderson and Barclay, 1995).

This work contributes to the knowledge base by presenting and quantifying the performance of three different methods used to detect the calls of the West Indian manatee. Within this paper the three detection algorithms are described and their performance is quantified by using an existing library of recorded manatee vocalizations. The results of this work may ultimately lead to the development of a manatee warning system that can warn boaters of the presence of manatees.

II. CHARACTERISTICS OF MANATEE VOCALIZATIONS

Other researchers have comprehensively described the distinctive characteristics of manatee vocalizations (Schevill and Watkins, 1965; Hartman, 1969; Steel, 1982; Bengston and Fitzgerald, 1985; Sousa-Lima *et al.*, 2002). Typical manatee vocalizations have a duration that lasts between 0.25 and 0.5 s. The vocalizations usually contain a fundamental frequency and several higher harmonics. For a majority of the manatee calls, the first harmonic dominates and contains a majority of the acoustic energy (see Fig. 2). The fundamental frequency ranges from 2 to 5 kHz, tends to be much smaller than the first harmonic, and is typically buried in background noise. The magnitude of the higher harmonics decreases as the frequency increases. As has been previously reported, the vocalizations can be variable and for some calls the fundamental frequency dominates over the first harmonic. This harmonic structure can be used to identify manatee vocalizations. Some of the ambient background noise that exists in the waters of interest results from boats, snapping shrimp, fish, birds, marine mammals, rain, wind, and wave motion. Any detection system needs to be able to discriminate manatee calls from background noise. It is assumed by the authors that the manatee vocalizations generated in the high-speed boat zones will have similar characteristics to the recordings already made by previous researchers.

III. DETECTION ALGORITHMS

A. Threshold method

The threshold method works by looking at the fast Fourier transform (FFT) of the input signal. The input signal is first filtered using a fourth-order Butterworth digital high-pass filter with a cutoff frequency of 1100 Hz. This frequency was chosen to filter out extraneous noise not associated with manatee vocalizations. The FFT of the signal is taken for each frame and its maximum value is determined. This maximum value is then compared to a user-defined threshold. If the maximum value of the FFT exceeds the threshold level, an output is produced to indicate that a manatee vocalization has been detected. A sample plot of the FFT of a manatee vocalization is shown in Fig. 2. The dashed vertical line indicates the cutoff frequency of the high-pass filter with the threshold detection region indicated by the cross-hatched region. Any signals that exceed the threshold level within the passband are considered to be manatee vocalizations. Although this method is simplistic, it

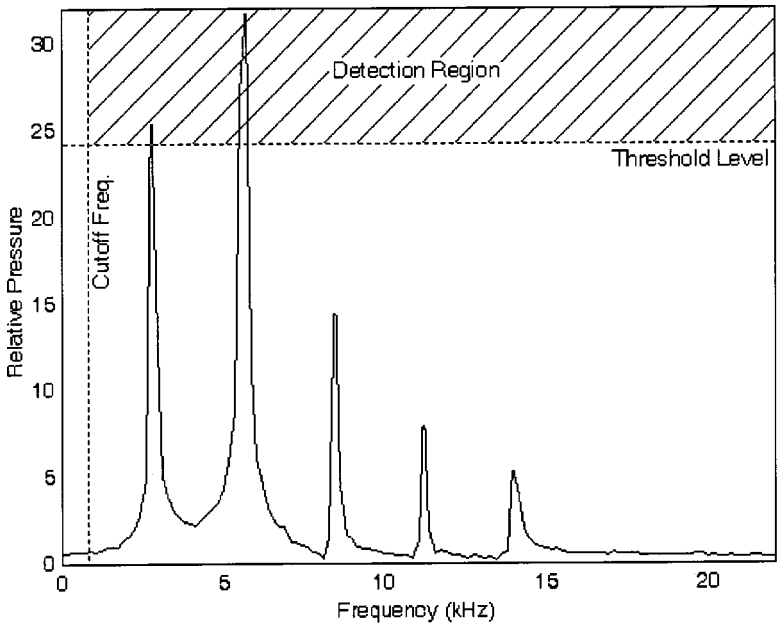


FIG. 2. Representation of a manatee vocalization and the threshold method.

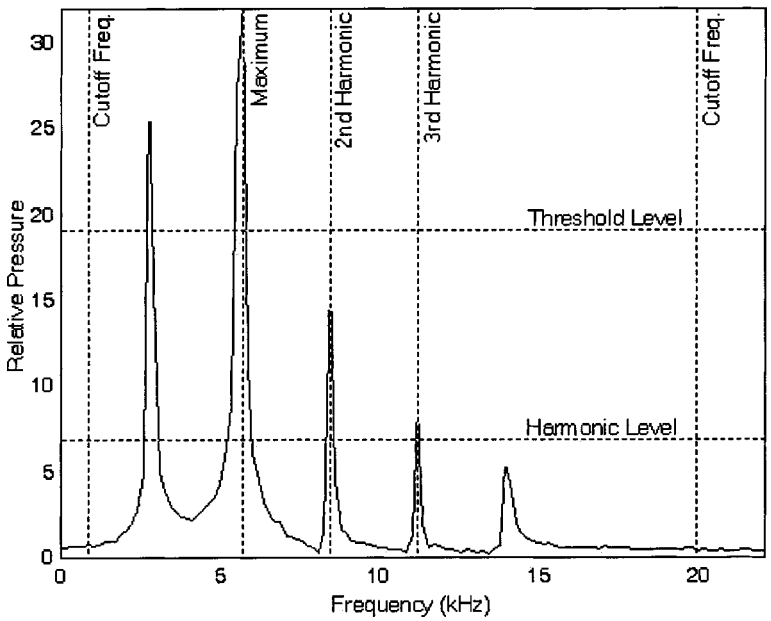


FIG. 3. Representation of a manatee vocalization and the harmonic content method.

is performed to provide a baseline for comparison to more sophisticated detection algorithms and to verify the need for a more complicated detection system.

B. Harmonic content method

The harmonic content method also works by looking at the input signal's fast Fourier transform. A filter is once again used to remove unwanted frequency components from the input signal. For this method, a Butterworth digital bandpass filter is chosen having a passband from 1100 to 20 000 Hz. Once the input signal is filtered, its FFT is performed and then several steps are taken to analyze the FFT of the signal to determine if a manatee vocalization is present. The first step is to compare the maximum value of the FFT to a specified threshold, as described in the threshold method. The next step is to examine the frequency at which the maximum value of the FFT occurs and look for the associated harmonics. The frequency of the maximum value is determined and then the FFT of the input signal is checked to determine if any harmonics are present at 1.5 times and 2 times the frequency of the maximum value. Analysis of manatee vocalizations has shown that the first harmonic of the vocalization often contains the most energy. If this is the case, the frequency of the maximum value needs to be multiplied by one and a half times in order to find the second harmonic and two times to find the third. If the maximum value occurs at the fundamental frequency of the manatee call, no peak will be present at the 1.5 times the fundamental frequency. Once the harmonic frequencies are determined, the values of the FFT at these frequencies are found and compared to a specified threshold for each harmonic. If either of the values at the harmonic frequencies exceeds the threshold, the signal is said to have met the technique's requirements.

If the input signal exceeds the maximum value threshold and either of the harmonics is present, a logical true output is produced. If not, a logical false output is produced. It was found that the typical manatee call lasts approximately 0.3 s, while most background noise with similar sound structures, such as snapping shrimp, only last approximately 0.1 s. In order to help distinguish between a manatee vocalization and the sounds of snapping shrimp, a final check is applied to the signal. If the duration of the logical true output does not remain true for a specified amount of time ($\approx$0.3 s), the signal is not considered to be a manatee vocalization. If the logical signal remains true for more than 0.3 s, the algorithm indicates that a manatee vocalization has been made. The harmonic content method is illustrated in Fig. 3. The dashed vertical lines represent the cutoff frequencies of the bandpass filter and the frequencies of the maximum value and its harmonics. The dashed horizontal lines show the specified values for the maximum and the harmonic value thresholds.

C. Autocorrelation method

A flowchart illustrating the steps is presented in Fig. 4. The analog input into the autocorrelation method is sampled at 44.1 kHz and then placed into a 128-sample buffer (3 ms). The signal is then filtered with a fourth order Butterworth bandpass filter with a high-pass cutoff frequency of 550 Hz and a low-pass cutoff frequency of 20 kHz. Much of the random noise in the signal is removed using the autocorrelation function. The autocorrelation minimizes random components of the signal and allows other components (i.e., harmonic content) to pass relatively unchanged. The unbiased autocorrelation, R_{xx}, of a vector input, x, is shown in Eq. (1) (Bendat and Piersol, 1986). The vector x consists of N data points sampled every Δt, where $x(t)=x(n\Delta t)$. The lag number is represented by r, and m represents the maximum lag number. For this method N was chosen to be 128 and the maximum lag was also used ($m=127$)

$$R_{xx}(r\Delta t)=\frac{1}{N-r}\sum_{n=1}^{N-r} x_n x_{n+r} \quad 0\leqslant r\leqslant m. \tag{1}$$

The signal is then passed through another fourth-order Butterworth high-pass filter ($f_c=2.2$ kHz) to reduce any low-frequency content from the autocorrelation function. A

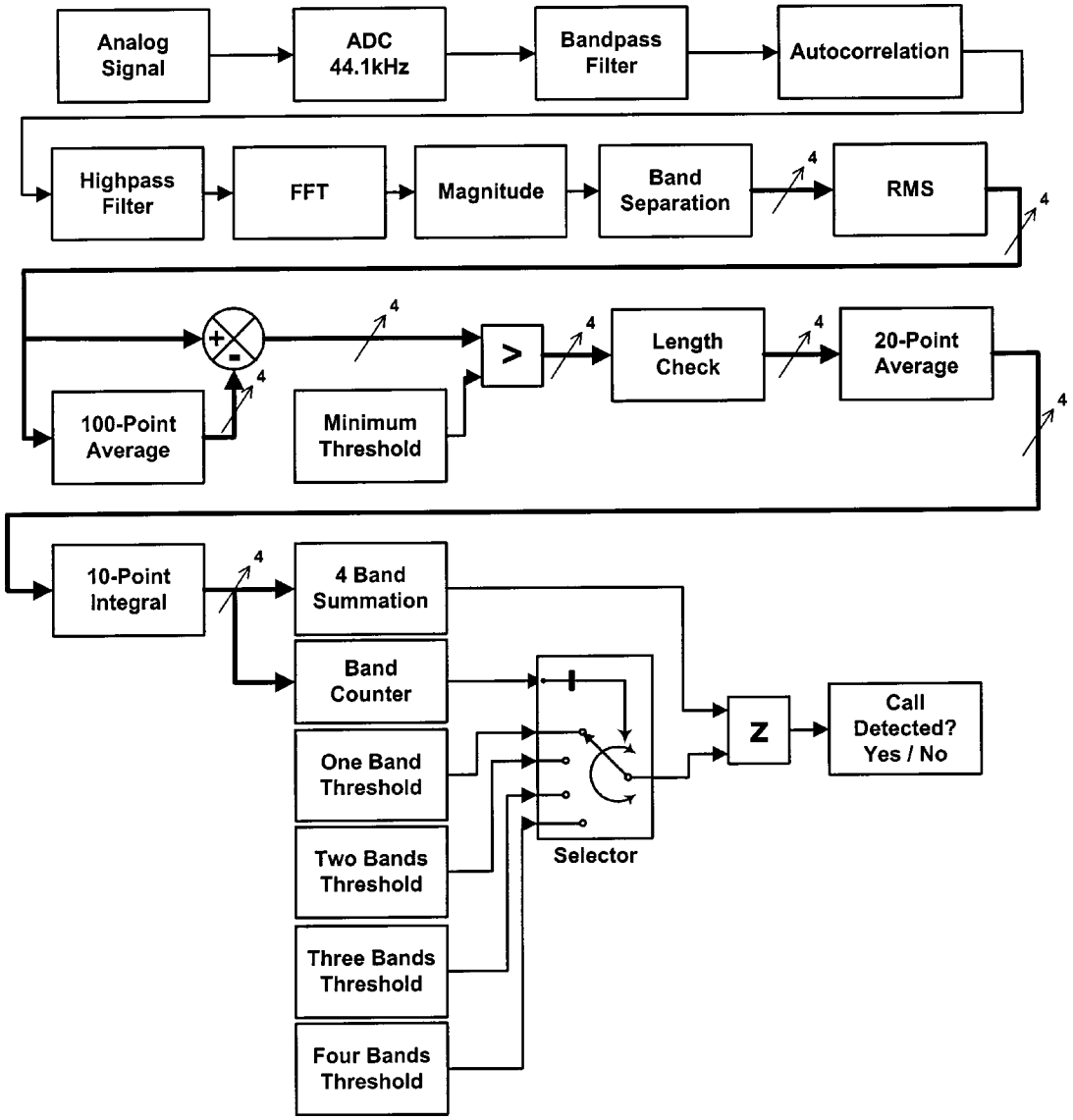


FIG. 4. Description of the autocorrelation method.

fast Fourier transform (FFT) is then taken, so that the frequency content of the autocorrelation signal can be analyzed. The magnitude of the complex FFT data is then calculated, using the 64 spectral lines. The FFT data are then split into four frequency bands ($0 \rightarrow 5$ kHz,$5 \rightarrow 10$ kHz,$10 \rightarrow 15$ kHz,$15 \rightarrow 20$ kHz), as referred to in the block called ''band separation.''

The root-mean-square of the four bands is taken and a 100-point moving average is subtracted from the current value to eliminate sustained sounds not associated with manatee vocalizations. The 100-point moving average covers approximately 0.3 s of data. Manatee vocalizations typically last between 0.25 and 0.5 s. A shorter moving average results in the subtraction of the manatees call from itself, and taking a longer moving average results in the algorithm not updating fast enough to subtract out shorter noises. Next, any sample that does not exceed a minimum threshold is set to zero. To help negate the effects of snapping shrimp, any sounds that do not exceed a specified time length are zeroed (''length check'' block). Because of the discontinuities that occur in the signal, a 20-point moving average is taken and a 10-point moving integral is used to create a continuous signal.

Four thresholds are used to determine if a manatee vocalization is present. A different threshold is selected based upon the number of frequency bands that contain nonzero data. The ''band counter'' block is used to count the number of bands that contain signals that are greater than zero. The output of this block is an integer value ranging from 0 to 4. This signal is used to specify which of the four ''band threshold'' blocks is used to compare with the output of the ''4-band summation'' block. The ''4-band summation'' block is used to add the signals for each of the four bands. This sum of the frequency bands is then compared with the appropriate threshold. If the sum of the amplitudes is greater than the selected threshold, then the algorithm indicates that a manatee vocalization has been detected.

IV. COMPARISON OF PERFORMANCE

The three detection algorithms were evaluated in the laboratory using prerecorded manatee vocalizations and

TABLE I. Description of sound recordings.

Track #	Comments	Length	# of Calls	SNR
1	Some shrimp snapping, no other major noises	2:05	17	4.51
2	Some shrimp snapping, no other major noises	2:05	16	4.56
3	Some shrimp snapping, no other major noises, quiet manatee calls	1:22	23	3.33
4	Constant shrimp snapping	2:03	26	4.68
5	Some shrimp snapping, engine noise	1:47	9	10.10
6	Some shrimp snapping, digestive sounds 4:20–4:40, engine noise at end	5:24	97	2.07
7	Overloaded hydrophones at the start, quiet manatee calls, no other major noises	2:08	38	2.56
8	Some shrimp snapping, bird-like calls, repetitive mechanical noise, hydrophone movement	5:08	64	2.58
9	Some shrimp snapping, no other major noises	1:51	44	2.80
10	Very little shrimp snapping, loud water noises at start of track, quiet recording level	5:15	85	2.63
11	Some shrimp snapping, repetitive mechanical noise, moving water sounds	3:04	40	1.58
12	Some shrimp snapping, repetitive mechanical noise, some bird-like calls, moving water sounds	3:01	31	1.40
13	Some shrimp snapping, repetitive mechanical noise	2:00	17	1.49
14	Some shrimp snapping, repetitive mechanical noise	0:34	5	2.25
15	Some shrimp snapping, bird-like calls	0:21	2	2.02
16	Some shrimp snapping, no other major noises	1:21	25	3.26
17	Some shrimp snapping, repetitive mechanical noise, some hydrophone movement, quiet manatee calls	5:37	67	2.45
18	Some shrimp snapping, some hydrophone movement, quiet manatee calls	0:42	9	2.26
19	Very little shrimp snapping, some moving water sounds	0:52	9	2.64
20	Very little shrimp snapping, moving water sounds, hydrophone movement	2:52	27	3.60

background noise from recordings made by O'Shea (O'Shea, 1981–1984). The original recordings were made using a Navy H-56 hydrophone and an Uher 4400 two-track tape recorder, operating at 19 cm/s (20–25-kHz bandwidth). The recordings were converted to a digital format on two compact disks. The tapes were digitized using a Soundblaster Live sound card in 48-KHz, 16-bit stereo using the Creative sound recorder. The three detection algorithms were tested

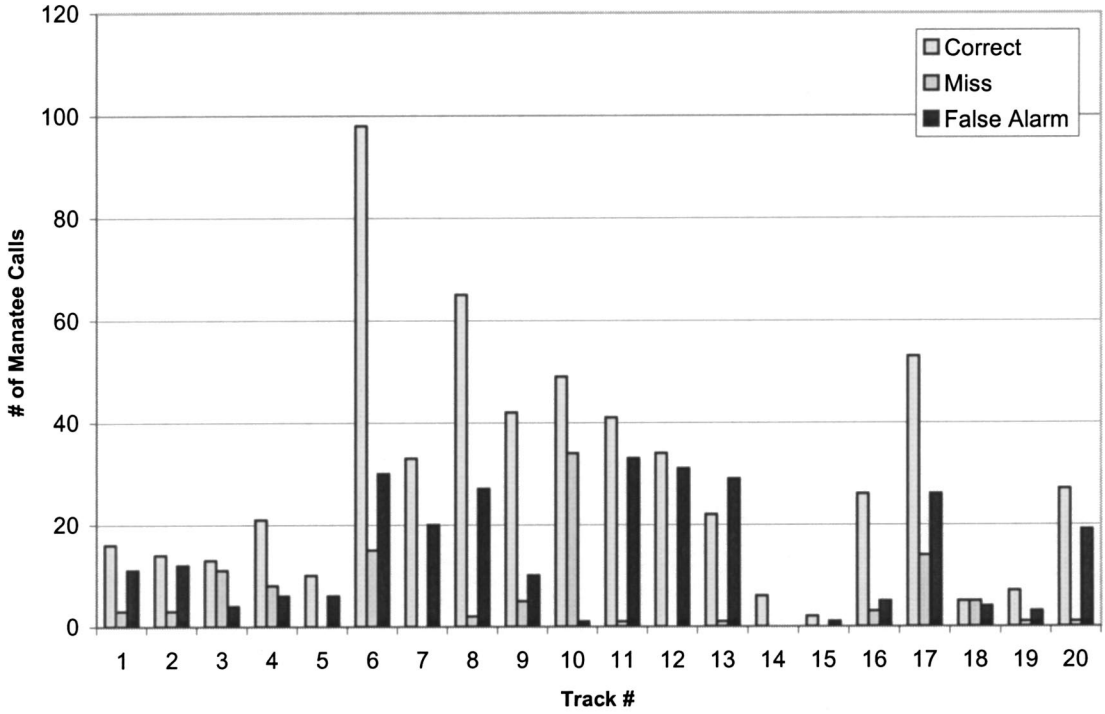


FIG. 5. Performance of the threshold method.

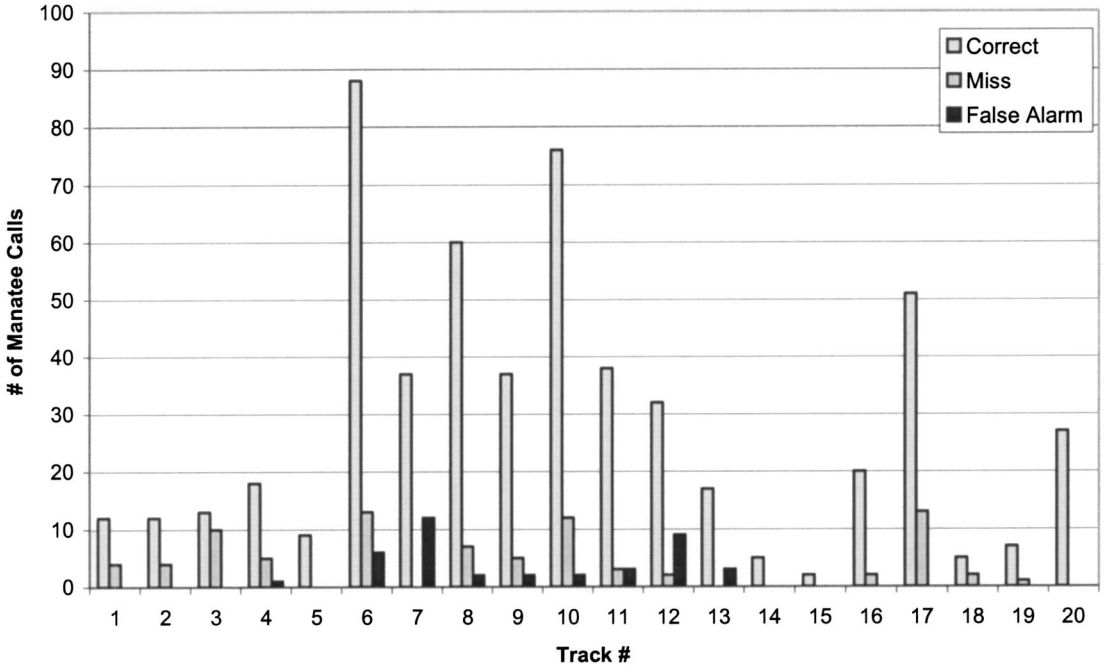


FIG. 6. Performance of the harmonic content method.

using the first 20 tracks of disk # 1. These 20 tracks contain 651 audible manatee vocalizations along with the sounds of snapping shrimp, boat engine noise, repetitive mechanical noise, bird-like calls, and water movement. A description of the sounds contained on each track is shown in Table I. Additionally, the signal-to-noise ratio (SNR) for each track is computed by comparing the ratio of the root-mean-square (rms) manatee vocalization signal to the rms background noise levels. The manatee vocalizations were spliced from the audio tracks for this calculation, resulting in a track containing only noise and a track containing only manatee vocalizations. For each of the 20 audio tracks, the rms values of signal and noise are computed in this manner.

The performance of each detection method is shown in Figs. 5–7 in terms of the number of calls correctly identified, the number of misses, and the number of false alarms. A ''miss'' refers to a situation in which a manatee vocalization was audibly heard but the system did not detect it. A false alarm refers to a situation in which a vocalization was not audibly heard and the system indicated that a manatee signal

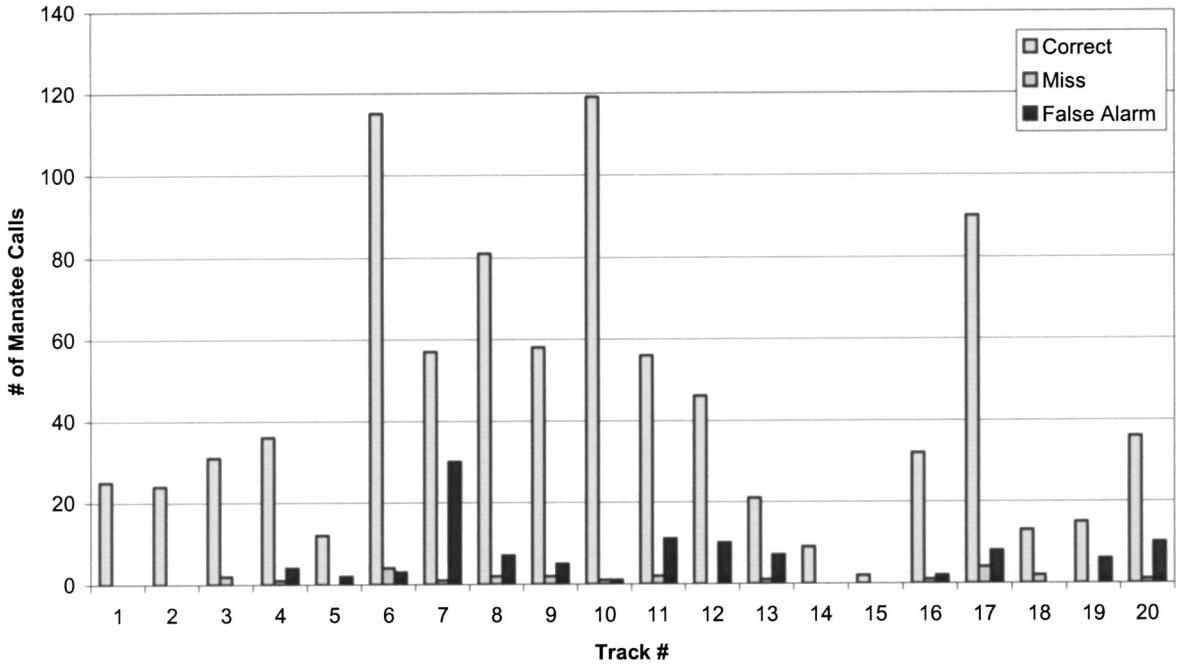


FIG. 7. Performance of the autocorrelation method.

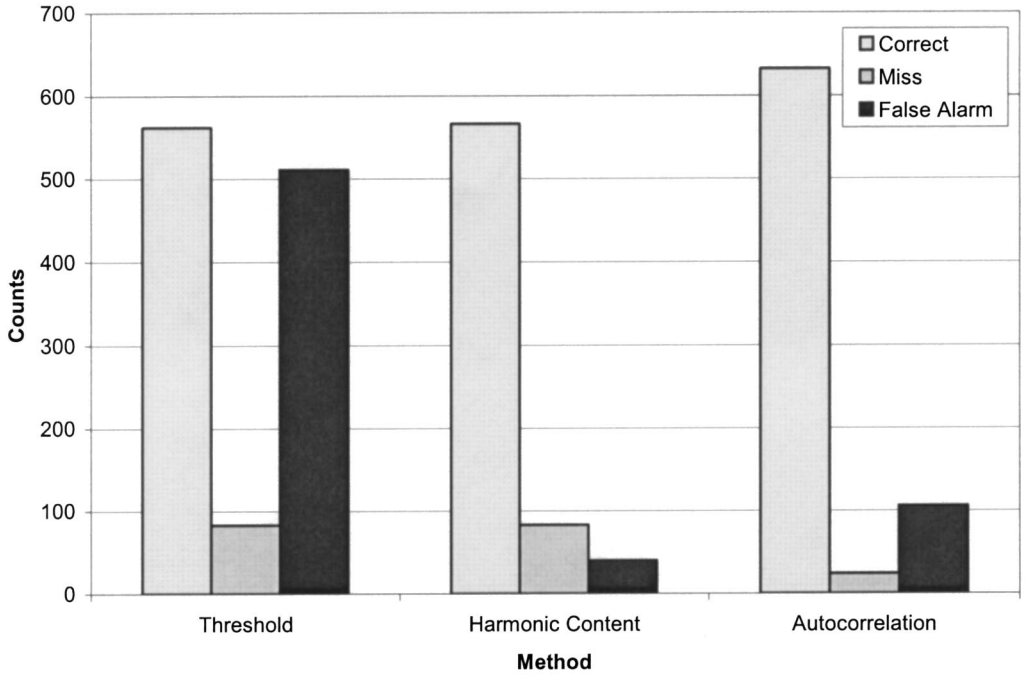


FIG. 8. Overall performance comparison for all detection algorithms.

was present. The overall comparison of the performance for the three detection systems is shown in Fig. 8 and listed in Table II. The performance of the threshold method detection system had approximately the same number of correct calls (87%) as it had false alarms (79%). This poor false alarm detection rate is attributed to the simplistic nature of the detection algorithm. It was important to implement this algorithm because it verified the hypothesis that a simple threshold detector, in the frequency range of interest, is not effective in detecting manatees. The harmonic content method identified approximately the same number of manatee calls correctly as the threshold method. However, it also resulted in a significantly reduced false alarm rate (6.1%). The autocorrelation method provided the best performance in terms of the number of manatee calls correctly identified (96%) and the number of misses (3.6%). The number of false alarms (16%) for the autocorrelation method was also significantly reduced compared to the threshold method.

For any detection system it is of primary importance to be able to detect a manatee if one is present in the vicinity of the detector. If a manatee is close to a detection system and the system does not provide a warning to an approaching boat, a collision may result. Therefore, the most important criterion is to make the number of misses as low as possible. Likewise, if the system perpetually warns boaters when there are no manatees present, the boaters will quickly learn to ignore the warning signals. Based on these criteria, the autocorrelation method provided the best overall performance of the three detection systems investigated.

V. CONCLUSIONS

Within this paper, three different detection algorithms are used to detect the calls of the West Indian manatee (threshold, harmonic content, and autocorrelation methods). The detection systems were tested in the laboratory using simulated manatee vocalizations from an audio compact disk. The autocorrelation method was able to achieve the overall best performance and was able to correctly identify ~96% of the manatee vocalizations. However, the system also resulted in a false alarm rate of ~16%. There are two important factors that will affect the performance of any manatee warning system: the signal strength of a manatee vocalization and the sound level of the background noise. Both of these factors are directly related to the signal-to-noise ratio of a manatee signal. Although the manatee recordings used within this work included background noise, neither of these factors was directly addressed. Ultimately, it is important to consider these issues in determining the range in which a manatee warning device will be effective. The next logical extension of this research is to investigate the manatee sound levels as a function of distance from the

TABLE II. Results of the three detection algorithms.

Method	Number correct	Number of misses	Number of false alarms	Percent correct	Percent missed	Percent false alarms
Threshold	562	83	511	87.13	12.87	79.22
Harmonic content	566	83	40	87.21	12.79	6.16
Autocorrelation	632	24	106	96.34	3.66	16.16

source. These factors will define the spacing of the hydrophones necessary to ensure that a boat can be slowed prior to entering a protected waterway. Another issue that needs to be considered is the vocalization rate and the distance a manatee and boat can travel within a period of vocalization. Although it has been previously observed that on average manatees vocalize at a rate of 1–5 times per 5-min period, more research needs to be performed on the duration of time that an individual manatee does not vocalize. Additionally, the existing detection system was not quantified using the full spectrum of sounds present in environment (i.e., fish and dolphin vocalizations). Prior to the implementation of an actual warning system, the system needs to be able to discriminate other biological sounds as well. If it is determined that the manatee signals have sufficient energy, the results of this work may ultimately lead to the development of a manatee warning system that can warn boaters of the presence of manatees.

ACKNOWLEDGMENT

The authors would like to express their sincere appreciation to the Florida Fish and Wildlife Conservation Commission in supporting this research.

Anderson, P. K., and Barclay, R. M. R. (**1995**). ''Acoustic signals of solitary dugongs: physical characteristics and behavioral correlates,'' J. Mammal. **76**(4), 1226–1237.

Bendat, J. S., and Piersol, A. G. (**1986**). *Random Data Analysis and Measurement Procedures* (Wiley, New York), pp. 385–391.

Bengston, J. L., and Fitzgerald, S. M. (**1985**). ''Potential Role of vocalizations in West Indian manatees,'' J. Mammal. **66**(4), 816–819.

Evans, W. E., and Herald, E. S. (**1970**). ''Underwater calls of a captive amazon manatee, *Trichechus inunguis*,'' J. Mammal. **51**(4), 820–823.

Gerstein, E. R. (**2002**). ''Manatees bioacoustics and boats,'' Am. Sci. **90**(2) (March–April), http://americanscientist.org/articles/02articles/gerstein.html

Gerstein, E. R., Gerstein, L., Forsythe, S. E., and Blue, J. E. (**1999**). ''The underwater audiogram of the West Indian manatee (*Trichechus manatus*),'' J. Acoust. Soc. Am. **105**(6), 3575–3583.

Hartman, D. S. (**1969**). Nat. Geogr. **136**, 342–353.

Nair, R. V., and Lal Mohan, R. S. (**1975**). ''Studies on the vocalization of the sea cow dugong in captivity,'' Indian J. Fisheries **22**, No. 1 & 2, 277–278.

Nowacek, D. P., Casper, B. M., Wells, R. S., Nowacek, S. M., and Mann, D. A. (**in press**). ''Intraspecific and Geographic Variation of West Indian Manatee,'' J. Acoust. Soc. Am.

O'Shea, T. (**1981–1984**). ''Manatee vocalization catalog of sounds,'' produced by: Coastal Systems Station, Naval Surface Warfare Center, Dahlgren Division, Panama City, FL.

Richard-Clark, E. M. (**1991**). ''A vocalization comparison of the West Indian manatee in Florida,'' M.S. thesis, Florida Institute of Technology, Melbourne, FL.

Schevill, W. E., and Watkins, W. A. (**1965**). ''Underwater calls of *Trichechus* (manatee),'' Nature (London) **205**, No. 4969, 373–374.

Sonoda, S., and Takemura, A. (**1973**). ''Underwater sounds of the manatees, *Trichechus manatus and T. inunguis (Trichechidae)*,'' Report for the Institute of Breeding Research, Tokyo University of Agriculture, No. 4, 19–24.

Sousa-Lima, R. S., Paglia, A. P., and Da Fonesca, G. A. B. (**2002**). ''Signature information and individual recognition in the isolation calls of Amazonian manatees, *Trichechus inunguis* (Mammalia: Sirenia),'' Anim. Behav. **63**, 301–310.

Steel, C. (**1982**). ''Vocalization patterns and corresponding behavior of the West Indian manatee (*Trichechus manatus*),'' Ph.D. dissertation, Florida Institute of Technology, Melbourne, FL.

Evaluation of an auditory model for echo delay accuracy in wideband biosonar

Mark I. Sanderson[a)]
Department of Neuroscience, Brown University, Providence, Rhode Island 02912

Nicola Neretti
Brain Sciences, Brown University, Providence, Rhode Island 02912

Nathan Intrator
Institute for Brain and Neural Systems, Brown University, Providence, Rhode Island 02912

James A. Simmons
Department of Neuroscience, Brown University, Providence, Rhode Island 02912

(Received 27 November 2002; revised 23 May 2003; accepted 16 June 2003)

In a psychophysical task with echoes that jitter in delay, big brown bats can detect changes as small as 10–20 ns at an echo signal-to-noise ratio of ~49 dB and 40 ns at ~36 dB. This performance is possible to achieve with ideal coherent processing of the wideband echoes, but it is widely assumed that the bat's peripheral auditory system is incapable of encoding signal waveforms to represent delay with the requisite precision or phase at ultrasonic frequencies. This assumption was examined by modeling inner-ear transduction with a bank of parallel bandpass filters followed by low-pass smoothing. Several versions of the filterbank model were tested to learn how the smoothing filters, which are the most critical parameter for controlling the coherence of the representation, affect replication of the bat's performance. When tested at a signal-to-noise ratio of 36 dB, the model achieved a delay acuity of 83 ns using a second-order smoothing filter with a cutoff frequency of 8 kHz. The same model achieved a delay acuity of 17 ns when tested with a signal-to-noise ratio of 50 dB. Jitter detection thresholds were an order of magnitude worse than the bat for fifth-order smoothing or for lower cutoff frequencies. Most surprising is that effectively coherent reception is possible with filter cutoff frequencies well below any of the ultrasonic frequencies contained in the bat's sonar sounds. The results suggest that only a modest rise in the frequency response of smoothing in the bat's inner ear can confer full phase sensitivity on subsequent processing and account for the bat's fine acuity or delay.
[DOI: 10.1121/1.1598195]

PACS numbers: 43.80.Lb, 43.64.Bt [WA]

I. INTRODUCTION

It is widely accepted from behavioral and physiological evidence that echolocating bats determine the distance to objects, or target range, from the time delay of frequency-modulated (FM) echoes (Grinnell, 1995; O'Neill, 1995; Schnitzler and Henson, 1980; Simmons, 1973, 1980, Simmons and Grinnell, 1988; Sullivan, 1982). In two-choice or yes–no discrimination tests, big brown bats (*Eptesicus fuscus*) can distinguish differences in echo delay as small as 50–100 μs (equivalent to 1–2 cm of target range; Moss and Schnitzler, 1995; Simmons and Grinnell, 1988). These behavioral thresholds are roughly consistent with the lower limits (~100–300 μs) for the accuracy of echo-delay registration by response latency in single neurons of the big brown bat's auditory midbrain, and the sharpness of delay tuning in individual forebrain neurons (Dear *et al.*, 1993; Feng *et al.*, 1978; O'Neill and Suga, 1982; Pollak *et al.*, 1977). However, big brown bats can detect much smaller changes in delay in a different behavioral procedure where echoes jitter in delay from one broadcast to the next (Moss and Schnitzler, 1995). In several versions of the jittering-echo experiment, conducted in two different laboratories, bats were able to detect changes as small as 0.5–1 μs (Menne *et al.*, 1989; Moss and Schnitzler, 1989; Moss and Simmons, 1993). The bat's threshold for jitter was measured to be 10–15 ns when an apparatus was developed to test even smaller delay jitter (Simmons *et al.*, 1990). At a controlled echo signal-to-noise ratio of 36 dB, big brown bats can detect delay changes as small as 40 ns, which actually is possible from an information-theoretic perspective if the bats used several successive broadcasts to judge whether echoes jittered in delay (Simmons *et al.*, 1990). Nevertheless, the degree of temporal precision required to support this performance has been described as virtually impossible for the auditory system to achieve, so the ''10-ns result'' is widely assumed to be due to an artifact, most likely spectral in nature, rather than perception of such small changes in time (Beedholm and Mohl, 1998; Menne *et al.*, 1989, Pollak, 1993; Schnitzler *et al.*, 1985). The work reported here concerns whether models of peripheral auditory transduction and coding are in fact unable to support submicrosecond delay perception at levels achieved by echolocating bats.

a)Electronic mail: mark_sanderson@brown.edu

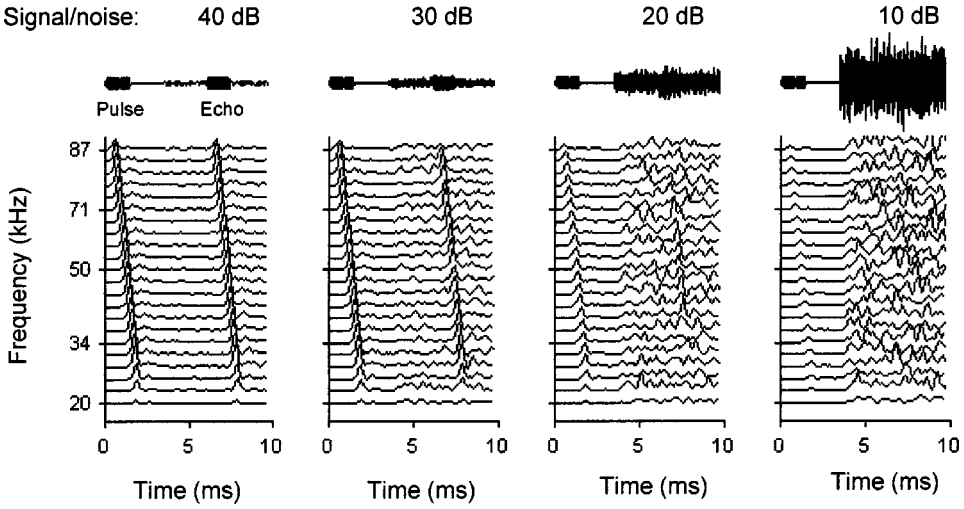


FIG. 1. Time waveforms and their filterbank output for different signal-to-noise ratios. A pulse (90–20 kHz, linear FM) was followed by an echo at 6-ms delay. The pulse and echo are identical, except that bandlimited noise (90–20 kHz) was added to the signal 2 ms after the pulse. The signal-to-noise ratio decreases until the echo is no longer visible, which in these examples occurs at 10 dB. The scale for the time waveforms at top is the same for all signal-to-noise ratios. The filterbank output for each waveform is plotted in the bottom row (see the text for details). The scaling for the four bottom plots varies to display the full data range in each plot.

This paper describes an auditory filterbank model of the bat's auditory periphery and the accuracy of this model for determining echo delay compared to standard sonar signal-processing techniques (cross correlation) as well as a previously used filterbank model of the bat's auditory periphery (Menne and Hackbarth, 1986; Hackbarth, 1986). We simulated the essential signal-processing steps performed by the bat's inner ear in multiple-trial tests to evaluate model performance for detecting jitter in echo delay. These tests allowed us to identify the values of model parameters required to account for the bat's submicrosecond acuity and the 10-ns result. One goal is to guide future physiological experiments by determining whether the bat's performance really is impossible, or whether there is a range of physiologically testable model parameters that makes this performance possible.

The question of the bat's delay acuity has been addressed previously for cross-correlation receivers (Menne and Hackbarth, 1986) and for a simple filterbank receivers (Hackbarth, 1986; Menne, 1988). Although fine delay acuity is possible with cross correlation, it cannot be achieved with the filterbank model as it was constituted originally. We chose to reexamine this question because several aspects of the Hackbarth (1986) filterbank model were unrealistic in auditory terms and several critical parameters have never been measured in bats. In particular, the design of the filterbank's smoothing (low-pass) filters removed all phase information from the envelopes to be processed for echo delay. Because several behavioral experiments have shown that bats may be able to detect changes in relative phase (Altes, 1984; Simmons *et al.*, 1990, Moss and Simmons, 1993), we decided to revisit the question of echo-delay accuracy and focus our attention on the role of the smoothing filter, which had been singled out as a critical stage for replicating the bat's performance (Simmons, 1980). A companion paper (Neretti *et al.*, 2003) addresses the related question of the delay resolution achievable by a similar auditory model. It should be clearly stated that *resolution* and *accuracy* refer to *separate* features of an echo-delay estimator, and the terms should not be used interchangeably (see Schnitzler and Henson, 1980). The accuracy of a delay estimator is the uncertainty in its estimate for the arrival time of echoes from a single reflecting point. In contrast, delay resolution refers to the minimum separation between two nearly simultaneous, overlapping echoes where the two reflections are assigned their own delay estimates instead of being interpreted as a single echo.

II. METHODS

A. Biosonar signals

The characteristics of the model are shown in the left panel of Fig. 1. The procedure for testing the model simulated the bat's biosonar emission, or pulse, in an acoustic environment that contained a single point-like reflector at a target range of ~1 m. For simplicity, the bat's pulse was modeled with a single harmonic, linear frequency-modulated (FM) chirp that swept from 90 to 20 kHz in a total duration of 1.5 ms. The test echo was a copy of the pulse shifted in time by an appropriate delay for a reflector at a range of 1.032 m (6 ms). The pulse and test echo were placed together in a single time record 12 ms long (top traces in Fig. 1).

This simulation assumed, as have other models, that the bat acquires a noise-free estimate, or template, of its outgoing pulse (see Menne and Hackbarth, 1986, Matsuo *et al.*, 2001, Saillant *et al.*, 1993). Bandlimited Gaussian noise (90–20 kHz) was added to the time record beginning 2 ms after the offset of the pulse and continued over the entire epoch in which the echo was embedded (Fig. 1, the 30-dB panel, shows an example of the onset of this noise). The echo signal-to-noise ratio was determined according to the following equation (Menne and Hackbarth, 1986):

$$\text{Signal-to-noise ratio(dB)} = 20\log_{10}(\sqrt{d}), \tag{1}$$

$$d = 2E/N_0, \tag{2}$$

where E is the echo energy flux, or the integral of squared amplitude over the duration of the signal, units are in pascals^{2}s; N_0 is the noise power per unit Hz, or mean-square amplitude divided by bandwidth, units are in pascals^{2}s.

Foraging bats in their natural habitats deal with acoustic environments that consist of more than just a single reflector and Gaussian noise. Instead, bats encounter non-Gaussian atmospheric effects and considerable clutter—other bats' emissions, multiple insects, background vegetation, etc. To

be clear, the noise added in these simulations was *not* meant to simulate noise in the bats' real-world acoustic environment. The added noise simulated only the controlled white noise the bats encountered in psychophysical experiments with jittering echoes (for details on the experimental signal-to-noise ratio, see Simmons *et al.*, 1990). Bats flying in their natural environments often deal with signal-to-noise ratios lower than the 30–50 dB. Nevertheless, our intention was to understand how a signal-processing model performed in a situation similar to that seen by the bat in experiments from which researchers have obtained precise, informative measures of their sonar capabilities.

The echo was just a single delayed reflected replica of the broadcast; Doppler shifts and those scattering properties of the target that affect the echo spectrum were not considered here (See Neretti *et al.*, 2003, for the resolution of a model that deals with multiple echoes.)

The pulse–echo time waveforms for four signal-to-noise ratios are shown in Fig. 1 (top row). For simplicity, the pulse and echo were set to have equal energy, while noise power increases from the left to the right in successive plots. Below the time traces in Fig. 1 are the 22 parallel filterbank outputs, which consist of envelopes that trace the FM sweep in the pulse and the echo in a spectrogram-like format. The location of the echo is obvious in the time waveform and filterbank output for high signal-to-noise ratios (e.g., signal-to-noise ratios 40, 30 dB in Fig. 1). At lower signal-to-noise ratios, however (e.g., 10 dB), the noise swamps the presence of the echo in both the waveform and filterbank displays. Our simulations of echo-delay determination assessed how different versions of the filterbank channels performed in varying levels of noise. Note that this model does not consider the effect that the receiver's internal noise has upon delay accuracy. That is, the filterbank itself and subsequent processing steps are assumed to be noise-free; only the external signal-to-noise ratio is included here.

B. Filterbank

We modeled the bat's cochlea with a filterbank composed of 22 channels, each of which had three components connected in series: a bandpass filter, a half-wave rectifier, and a low-pass filter. The frequency tuning of the ''cochlea'' was implemented with a series of overlapping bandpass filters (Chebyshev IIR filters, constant 4-kHz bandwidth). We chose the Chebyshev design because it allows a narrower bandwidth than the Butterworth design of the same order. Simulation of the transduction and capacitance of the inner hair cell was implemented with a half-wave rectifier and low-pass filter (Chebyshev IIR filters). The resulting output from the low-pass filter corresponded to the probability of neurotransmitter release at the base of the inner hair cell and represented the excitation function delivered to auditory-nerve fibers for coding as spikes. The model's filterbank parameters are listed in Table I. The center frequencies (CFs) for the bandpass filters were chosen to (1) cover the 20–90-kHz frequency range of the big brown bat's hearing (Koay *et al.*, 1997), and (2) to overlap as closely as possible with their neighbors at their 3-dB down points.

TABLE I. Values for various parameters of the filterbank model and delay estimation methods.

Sample rate	2 MHz
Signal	
Duration	1.5 ms
Sweep type	Linear
Harmonics	1
Frequency start	90 kHz
Frequency end	20 kHz
Rise/fall time	0.15 ms, cosine^2 ramp
Bandpass filter	
Number of filters	22; see Table II
Design	Chebyshev type 1 (IIR)
Ripple in passband	6 dB
Order	8
Bandwidth$_{3dB}$	4 kHz
Nonlinearity	
Half-wave rect. $y=x$ for $x\geqslant 0$	
$y=0$ for $x<0$	
Low-pass filter	
Design	Chebyshev type 1 (IIR)
Order	2, 5
Corner$_{3dB}$ freq.	1, 2, 4, 8, 125 kHz
Other	
A priori window	± 1000 μs

Filter center frequencies were restricted to integer multiples of the 0.5-μs sample period (listed in Table II). This step was to minimize interference between the simulation's sample rate (2 MHz) and filter CF. This restriction provided better digital approximations of the equivalent analog filter impulse responses. Otherwise, if this step was not taken, the outputs of filters with certain CFs exhibited interference effects with the sample rate in the peak region of the impulse

TABLE II. Period and center frequency of the 22 bandpass filters used in the filterbank.

	μs	kHz
1	50	20.000 000 0
2	44	22.727 273 0
3	39.5	25.316 456 0
4	35.5	28.169 014 0
5	32.5	30.769 231 0
6	29.5	33.898 305 0
7	27	37.037 037 0
8	25	40.000 000 0
9	23	43.478 261 0
10	21.5	46.511 628 0
11	20	50.000 000 0
12	19	52.631 579 0
13	18	55.555 556 0
14	17	58.823 529 0
15	16	62.500 000 0
16	15	66.666 667 0
17	14	71.428 571 0
18	13.5	74.074 074 0
19	13	76.923 077 0
20	12.5	80.000 000 0
21	12	83.333 333 0
22	11.5	86.956 522 0

response envelope. If these filters were used to estimate delay the resulting estimates would be biased away from the true peak in the impulse response. Restricting the filter CFs to those in Table II ensured that the digitally implemented impulse responses for all filters had the same envelopes. This minimized any potential bias in the delay estimates in individual channels.

The primary question was how the filterbank receiver's delay accuracy compares against a cross-correlation receiver. Here, we adopted a similar trial-by-trial approach to that used previously (Menne and Hackbarth, 1986; Hackbarth, 1986). The task was to estimate the delay between a pulse and a single echo at a fixed delay for multiple occurrences of the pulse and the echo. We implemented four delay estimation methods for comparison purposes.

C. Delay estimation methods

Three methods (''analog sum,'' ''Saillant,'' and ''Hackbarth'' methods) were used to estimate the delay between events corresponding to the pulse and echo generated by the filterbank. Our interest was focused on how the filterbank design parameters, not the different delay estimation methods, affected delay estimation accuracy. The estimation methods described below were not intended to realistically model the full range of physiological processes that the bat auditory system uses to estimate pulse–echo delay subsequent to auditory transduction by the inner ear. Instead, the aim was to measure, with the best precision possible, the temporal intervals between events in the filterbank channels.

1. Cross correlation

To obtain an optimal (matched-filter) delay estimate, we first computed the cross-correlation function between the noise-free pulse and the echo embedded in noise for each noise iteration at each signal-to-noise ratio. Then, the peak in the cross-correlation function was located within an *a priori* window ± 1000-μs window around the true delay (as in Menne and Hackbarth, 1986 and Hackbarth, 1986). Finally, the three points around the peak were fit with a quadratic function in order to determine the peak's precise position. By this interpolation procedure, delay estimates could be ''recorded'' with a precision greater than the 0.5-μs sample period we used to generate our pulse and echo signals.

The time-estimation process was repeated for each of the 400 trials with independent noise samples in a Monte Carlo procedure (see below). The delay estimate for each trial can be compared against the theoretical accuracy possible for a cross-correlation receiver (Burdic, 1968)

$$\sigma = (2\pi B d)^{-1}, \tag{3}$$

where d is from Eq. (2), and B is the sonar emission's noncentralized root-mean-square bandwidth, which for this simulation equaled 57.79 kHz. This value was slightly higher than the typical value, ~55 kHz, estimated for signals recorded in psychophysical experiments (Simmons *et al.*, 1990). The cross-correlation approach did not use the output from the filterbank, whereas the next three methods did.

2. Analog sum method

The second method for estimating delay, which will be called the ''analog sum'' approach, used the ''cochlear'' filterbank to process broadcasts and echoes. First, the output of all the filter channels was dechirped to remove the slope of the FM sweep by aligning the peaks of the outputs for the emitted pulses. Peaks for the broadcast in individual channels were located by finding the highest value in the channel output over the time window containing the broadcast [see Fig. 2(C)] and sliding each time series signal by a corresponding amount to align the peaks in different channels to the same reference time value. Then, the dechirped output was summed across all frequency channels [Fig. 2(B)], and the position of the peak in the resulting sum was located within an *a priori* window ± 1000 μs around the true delay. Finally, using a quadratic fit, the interpolated peak location was determined and stored as the overall estimate for echo delay. Thus, as for cross correlation, delay estimates have a higher precision than the 0.5-μs sampling interval. This process was repeated for each of the 400 trials with independent noise samples in a Monte Carlo procedure (see below).

3. Hackbarth method: Single peak detection followed by interval measurement

This approach, described by Hackbarth (1986), also used the filterbank to process broadcasts and echoes. First, the location of the peak in the output from each filter channel for the noise-free pulse was located inside a window from time zero to 3.5 ms (Fig. 1). Then, the peak for the echo was located inside the *a priori* window of ± 1000 μs centered on 6 ms after the peak for the broadcast [Fig. 2(C)]. Following this initial procedure, the method produced single triggered events for the pulse and for the echo. The method

(a) fit the three points around the pulse and echo peak samples with quadratic functions in order to determine their exact positions, which were marked by single pulses [this was performed separately for the pulse and echo response in each frequency channel, Figs. 2(D) and (E)];
(b) measured first-order intervals by subtracting the time of the interpolated pulse peak from the interpolated echo peak in each frequency channel [Fig. 2(F)];
(c) generated a point process by projecting these intervals across filterbank channels onto a single time axis [Fig. 2(G)];
(d) estimated the density of the point process along the time axis by convolving the points with a Gaussian kernel ($\sigma = 1$ μs, shift step $= 0.25$ μs [Fig. 2(H)]; and
(e) located the peak of the density function resulting from convolution by fitting a quadratic to the peak sample and its two neighboring samples, as before, and stored this interpolated value as the *overall delay estimate* for that trial.

This process was repeated for each of the 400 trials with independent noise samples in a Monte Carlo procedure (see below).

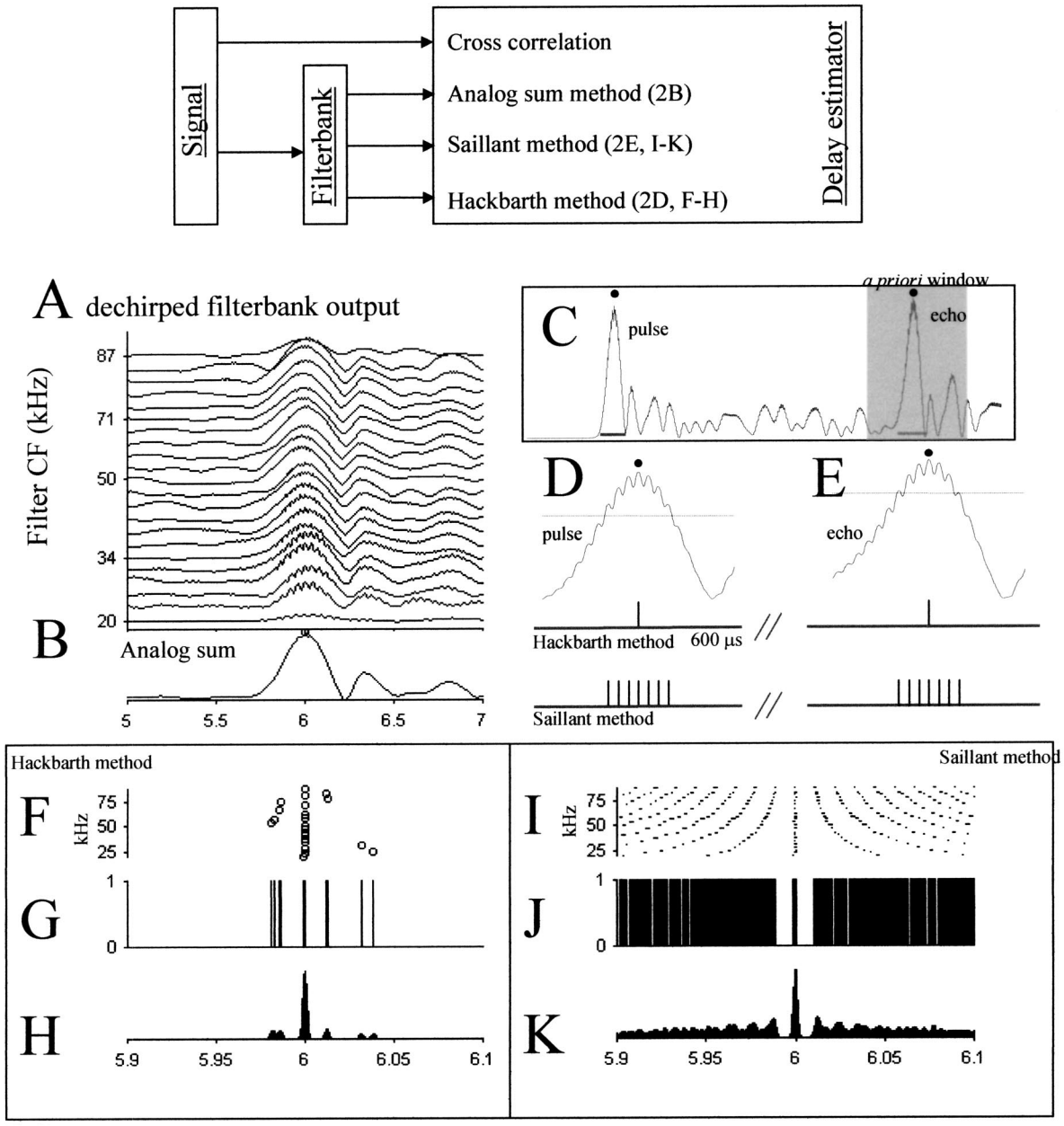


FIG. 2. Three methods for estimating delay from the filterbank output. The schematic at top illustrates how the biosonar signal passes through the filterbank, the output of which is processed by three different estimators. The cross-correlation estimator operates directly on the pulse–echo signals. (A) Dechirped output from a filterbank with a second-order low-pass filter cutoff set to 8 kHz, and a signal-to-noise ratio of 36 dB. (B) The summed output, across frequency channels, from panel (A). The peak position is the ''analog sum'' method's echo delay estimate. (C) Output from the filterbank's third channel (CF $=25.3$ kHz). (D) Expanded view of the pulse [corresponds to left gray bar in panel (C)]. Spikes are triggered by the maximum peak for the Hackbarth method and for the local peaks above a threshold for the Saillant method. (E) Expanded view of the echo with triggered spikes for both methods at bottom. (F) Interpolated delay estimates from the Hackbarth method. (G) Histogram of peaks from (F). (H) Smoothed histogram from (G), using Gaussian kernel with $\sigma=1$ μs. The interpolated peak is the Hackbarth method's final delay estimate. (I–K) Same as (F)–(H) but for the Saillant method. Note that x axes for (F)–(K) range from ±100 μs, whereas (A) and (B) range from ±1000 μs relative to actual echo delay.

4. Saillant method: Detection of multiple peaks followed by all-order interval measurement

This final method created a series of triggered spike events for the pulse and then for the echo using a procedure similar to the ''peak-detection'' approach of Saillant *et al.* (1993). First, the peak position in the noise-free window for the pulse in each filterbank channel was located (as above), and the corresponding sample times across channels were used to dechirp the filterbank output (each channel's time axis was realigned so the peak pulse time corresponded to time zero). Then, the model generated spikes within two *a priori* windows: one centered on the pulse and one on the echo. The pulse *a priori* window spanned the noise-free time window from time zero to 3.5 ms (Fig. 1). As before, the echo *a priori* window spanned ±1000 μs and was centered in time at the true echo delay.

For each filterbank channel, separate thresholds were set for the pulse and echo to define the time windows inside

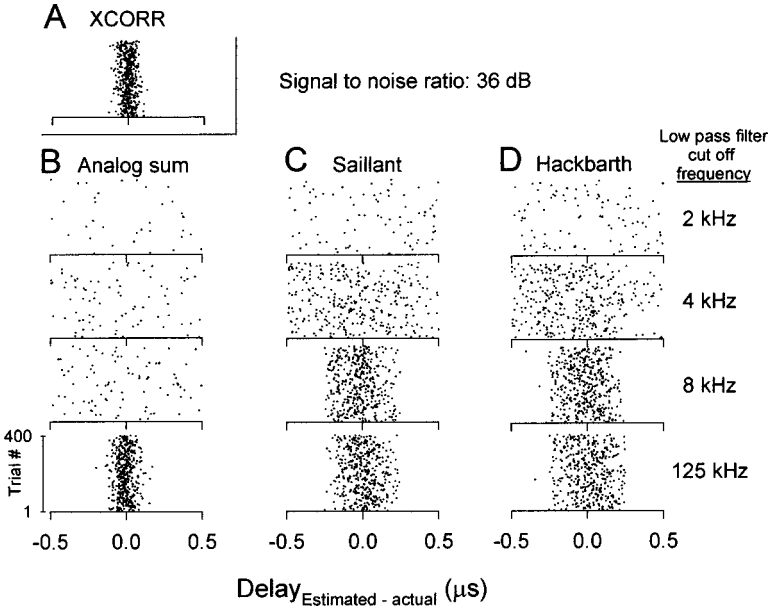


FIG. 3. Results from 400 Monte Carlo trials with echo signal-to-noise ratio fixed at 36 dB. (A) The cross-correlation receiver's estimates. (B), (C), and (D) Results for each of the three filterbank models are plotted in separate columns. As the low-pass filter cutoff frequency increased, the variability of the delay estimates decreased (1-kHz results not shown). Because the *x* axis is scaled so that only the central 1 μs of the full $\pm$ 1000-μs *a priori* window is visible, some of the estimates are not visible (especially for the 1- and 2-kHz low-pass filter conditions). The accuracy of each method was estimated by taking $\frac{1}{2}$ of the 68th percentile of the distribution of delay estimates, and is summarized in Fig. 4.

which multiple peaks in the filter output would registered as corresponding multiple spikes. For the pulse, the spike-window threshold was set at the value of 67% of the peak amplitude within the previously defined pulse *a priori* window. For the echo, the first step was to establish a noise threshold located 2 standard deviations above the mean noise level in the filter output over a time window containing just noise. In the second step, the echo spike-window threshold was set at the value 2/3 of the way between this noise threshold and the amplitude of the largest peak within the echo *a priori* window. Spike events then were generated for every local peak above the pulse-window and echo-window thresholds [see Figs. 2(D) and (E), respectively], and the location of each local peak was identified using interpolation (as before, a quadratic was fit to each local peak and the immediate neighboring samples). At this juncture, the channel-by-channel filterbank output is converted into two sets of spike events corresponding to all the local peaks within the pulse and echo windows [one channel's output shown in bottom panel of Figs. 2(D) and (E)].

To obtain a single estimate for pulse–echo delay from this multiple-spike representation, the method

(a) calculated the all-order intervals (Cariani and Delgutte, 1996) for pulse versus echo spikes within each channel [Fig. 2(I)];
(b) generated a point process by projecting these resulting intervals onto a single time axis [Fig. 2(J)];
(c) estimated the density of the point process along the time axis by convolution with a Gaussian kernel [$\sigma = 1$ μs, shift step$=0.25$ μs; Fig. 2(K)]; and
(d) located the peak of the resulting density function by fitting a quadratic to the peak sample and its two neighboring samples and stored it as the *overall delay estimate* for that trial.

This process was repeated for each of the 400 trials with independent noise samples in a Monte Carlo procedure (see below).

D. Monte Carlo trials

Our goal was to examine how the use of different filterbank parameters affected the variance of the delay estimation procedure. To estimate this variance we adopted a Monte Carlo procedure, a method that uses many independent trials, each with a different noise instantiation, in order to build a distribution of the delay estimates. The variance of the delay estimation procedure is then measured from this distribution. The number of trials used in the Monte Carlo simulations do not improve the accuracy of the model because there is no memory from trial to trial in the Monte Carlo method. Our results were identical using 100 or 400 trials. It follows that the number of Monte Carlo trials bear no relation to how many emissions the bats actually use in a single trial of the jitter task.

Pulse–echo delay estimation was repeated for 400 different realizations of noise added to the echo at a fixed signal-to-noise ratio [Fig. 3]. The true echo delay, 6 ms, was subtracted from each of the 400 estimates to form the error distribution. To estimate the mean and variance of the accuracy of this distribution we used a bootstrap procedure. Four hundred samples were drawn, with replacement, from the error distribution, and the 68th percentile was calculated. The resulting value was divided by 2 in order to match the standard deviation for a uniform distribution in a 1000-ms window ($\sim$683 μs). This sampling was then repeated 128 times to compute the bootstrapped estimate for the mean and standard deviation of accuracy.

The bootstrapped accuracy estimates are plotted in Fig. 4 against the range of signal-to-noise ratios tested. This method for estimating the estimates' variability was used for comparison with other models (Hackbarth, 1986, Menne and Hackbarth, 1986) and the analytical standard deviation calculation [see Eq. (3) above].

E. Additional constraints

The width (σ) of the Gaussian kernel chosen for the smoothing procedure affected the estimate of filterbank ac-

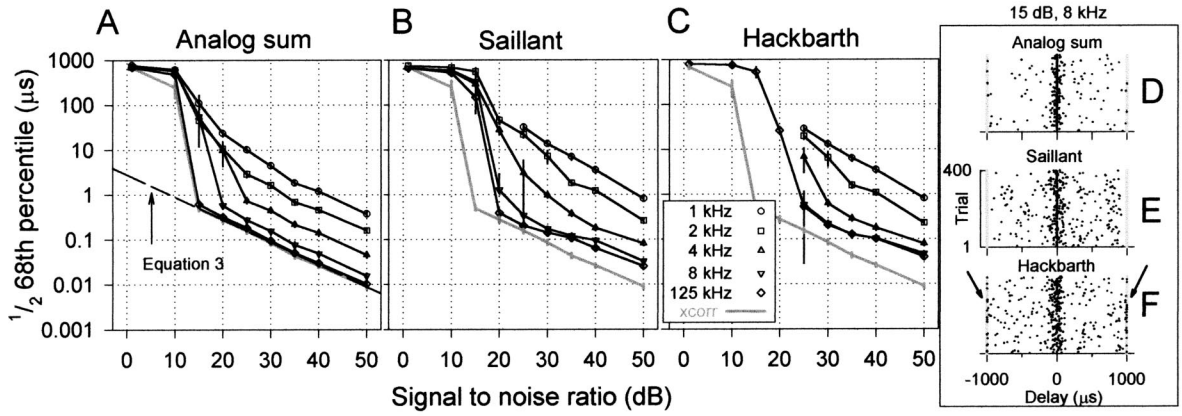


FIG. 4. Filterbank accuracy with different low-pass filter settings. The accuracy of the cross-correlation Monte Carlo simulations is plotted as a heavy gray line on each panel for comparison. (A) The accuracy of the analog sum method when applied to the filterbank output. Five different settings for the filterbank's low-pass filter were used [see the legend in panel (C)]. Increasing the low-pass filter cutoff value shifted the analog sum method's accuracy much closer to that of the cross correlation. The values for Eq. (3) are plotted as a dashed line. (B) Same as (A) except delay estimates were generated by applying the Saillant method to filterbank output. (C) Same as (A) except delay estimates were generated applying the Hackbarth method to filterbank output. (D)–(F) These three plots show the 400 trial-by-trial estimates for the analog sum, Saillant, and Hackbarth methods. Note that for the Hackbarth method, many estimates occur at the edges of the *a priori* window.

curacy (step *e* of the Hackbarth method and step *d* of the Saillant method). If the Gaussian was too wide relative to the ''true'' accuracy of the filterbank, all of the trial-by-trial estimates fell within a single bin. This underestimated the effective accuracy of the filterbank. On the other hand, if the Gaussian width was too narrow, the smoothing step resulted in multiple local peaks with equal heights and so failed to yield a single delay estimate. After testing of sample data with several different values for σ of the Gaussian kernel, we chose the value that yielded the maximum accuracy ($\sigma = 1$ microsecond).

The size of the *a priori* window also has a significant effect on the accuracy of each receiver design, as shown previously in Menne and Hackbarth (1986) and Hackbarth (1986). However, the *a priori* window size only affects the accuracy of the results within a fixed range of signal-to-noise values less than 15 dB, where the accuracy curve ''breaks'' or undergoes an abrupt decline due to ambiguity effects caused by the emergence of prominent sidelobes (Menne and Hackbarth, 1986). We chose the value of +/−1000 microseconds for our *a priori* window in order to compare our results with theirs. The use of *a priori* windows with respect to biosonar experiments is not unreasonable: the bat certainly knows when it vocalized its pulse, and in behavioral paradigms learns fairly quickly that most echoes return within a fixed time window.

F. Simulation of the jittering-echo experiment

In the jittering-echo experiment, the bat is trained to sit on a small platform and emit sonar sounds into microphones, and its task is to determine which of two loudspeakers returned echoes of those sounds that alternated in delay from one broadcast to the next (Simmons *et al.*, 1990). In individual jitter trials, the bat necessarily emitted at least two sounds to the jittering stimulus channel and two sounds to the nonjittering channel. A single stimulus echo was returned for each emission, the delay of which was fixed on one side and jittered between two values on the other side. The bat estimated each returning echo's delay for comparison with the delay of the next echo to then choose which side had the jittering echoes.

We simulated this experiment, following the method of Menne and Hackbarth (1986), in which the virtual bat discriminated a jittering target from a nonjittering target. On one ''side'' the virtual bat received two echoes embedded in noise at a fixed delay. On the other ''side'' the virtual bat received two echoes in noise with a temporal offset, or jitter, added to each echo. The data for the experiment had in fact been simulated under different conditions in the Monte Carlo simulations. We could therefore simply draw four delay estimates, without replacement, from these simulations and a temporal offset $(+\Delta t, -\Delta t)$ was added to the two delay estimates on the jittering ''side.'' On a single trial the virtual bat had to decide which side had the jittering echoes. The jitter experiment was simulated with 100 such trials, and the results of the simulation were expressed as the percentage of correct decisions for those 100 trials for different values of Δt (the amount of jitter, Δt, was varied from 0 to 20 microseconds). This entire procedure was then repeated 128 times to compute a bootstrap estimate of the variability for these percent-correct values.

III. RESULTS

A. Monte Carlo results

The results consist of a series of estimates of delay accuracy obtained with each of the four model designs on 400 Monte Carlo trials at each signal-to-noise ratio. To illustrate the nature of the simulation results, the 400 trial-by-trial delay estimates obtained with the four methods are plotted for one signal-to-noise ratio (36 dB) and different low-pass smoothing frequencies (1, 2, 4, 8, and 125 kHz) in Fig. 3. To establish a baseline for comparison, the performance of the cross-correlation procedure is show in Fig. 3(A), while the results from the three filterbank methods are plotted in sepa-

rate columns [Figs. 3(B)–(D)]. Each point on one of the graphs represents a single delay estimate for a pulse–echo pair with one iteration of independent noise added to the echo. All four methods generated well-behaved delay estimates across the Monte Carlo trials, and in each case the variability of the estimates changed with the low-pass cutoff. With increasing low-pass cutoff frequency, the variability of each distribution decreases appreciably (except for the Hackbarth method, for which the results at 8 and 125 kHz were very similar—that is, once the low-pass cuttoff was as high as 8 kHz, no further improvement in accuracy could be obtained). The 125-kHz low-pass filter condition was explicitly included in the simulations to observe what happened when all the available phase information was allowed to pass through the model and made available for processing (this was our method for the effective removal of the low-pass filter). As such, the analog sum method with a 125-kHz low-pass filter yielded the tightest distribution of delay estimates, a distribution that was indistinguishable from the results for the optimal delay estimator, the cross-correlation receiver, plotted in Fig. 3(A).

The summary and quantification of the variability of all delay estimates across a wide range of signal-to-noise ratios is plotted in Fig. 4. These curves show the effects of the low-pass filter parameters on the accuracy of delay estimates, and allow comparison of the performance for different estimation methods.

1. Cross-correlation accuracy

Because the cross-correlation receiver's accuracy was the lower bound to be expected for the various estimation methods, it is shown on each plot in Fig. 4 as a baseline. The cross-correlation results from the Monte Carlo trials fit the theoretical accuracy [Eq. (3), dashed line] for high signal-to-noise ratios ($>$15 dB). Between signal-to-noise ratios of 15 to 10 dB, the cross-correlation receiver accuracy falls off sharply compared to what theory predicts. The nature of this break in the cross-correlation performance was explored previously by Menne and Hackbarth (1986). They showed that Eq. (3) was applicable only when the signal noise ratio was above 15 dB, and our results are the same.

2. Filterbank model accuracy: Analog sum method

The analog sum method yielded results almost identical to the cross correlation when the entire signal bandwidth passed the low-pass filter (i.e., cutoff frequency$=$125 kHz). This shows that bandpass filtering, rectification, and peak picking had no deleterious effects on accuracy. However, as high-frequency phase information was progressively removed from the filterbank output by the low-pass filter, accuracy degraded precipitously. In the linear region of the results (signal-to-noise ratio $\geqslant$30 dB), the accuracy of delay estimation degraded by 2 orders of magnitude as the low-pass cutoff frequency was decreased from 125 to 1 kHz. Between signal-to-noise ratios of 20 and 15 dB, the accuracy curves break sharply from the linear case and eventually converge upon the standard deviation for the whole *a priori* window, 680 μs.

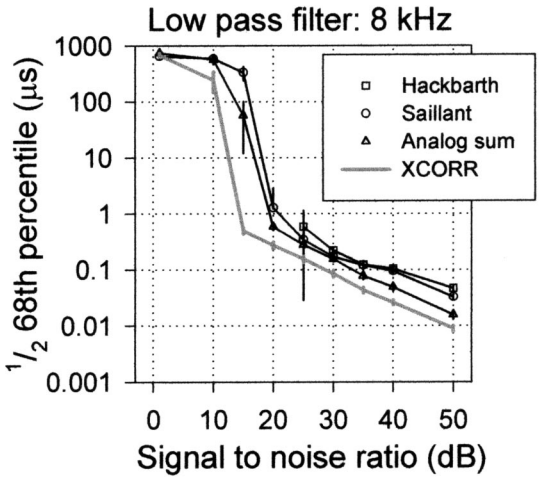


FIG. 5. Comparison of different filterbank delay estimation methods when using a second-order low-pass filter set at 8 kHz. Of the filterbank estimators, the analog sum method provided the best performance. There was an orderly arrangement to where each method ''broke,'' or underwent a sharp transition in accuracy: the cross-correlation accuracy broke between 15 to 10 dB, the analog sum method between 20 and 15 dB, the Saillant method between 25 and 20 dB, and the Hackbarth method between 30 and 25 dB. The Hackbarth and Saillant methods were identical for signal-to-noise ratios $\geqslant$30 dB.

3. Filterbank model accuracy: Saillant method

The Saillant method used one or more spikes to mark the occurrence of the pulse and echo in each channel [Figs. 2(D) and (E)]. As such, it uses less of the filterbank output's wave structure, compared to the analog sum method, to estimate pulse–echo delay and so would be expected to do worse at any given signal-to-noise ratio. In Fig. 4, the performance of the Saillant method is displaced upward relative to the analog sum method to slightly worse performance across most signal-to-noise ratios. This is more clearly shown in Fig. 5, which plots the performance of all four methods at an 8-kHz low-pass cutoff.

4. Filterbank model accuracy: Hackbarth method

In each frequency channel the Hackbarth method used only one spike to mark the occurrence of the pulse and one spike for the echo (see Fig. 2). At signal-to-noise ratios $\geqslant$30 dB, the accuracy results for the Hackbarth method were very similar to those of the Saillant method. The exception was the 125-kHz low-pass filter condition, which for the Hackbarth method did not show any appreciable difference from its 8-kHz result.

Unlike the Saillant method, we observed that for signal-to-noise ratios below 25 dB the *a priori* window size interacted with the low-frequency nature of the filterbank output. The width or duration of the ''excitation'' in each frequency channel was a function of the signal sweep rate, integration time, and low-pass filter time constant. Because the excitation half-width was about 250–300 μs [Fig. 2(D)], the outputs of the filters were such slowly varying signals that they could almost be considered as linear when viewed within a small *a priori* window. Consequently, the signals often reached their maximum at one of the window edges [Figs. 3(D), (E), (F)]. Because of this edge effect, we discarded any

accuracy measurements if a conspicuous number (>10%) of trials resulted in delay estimates at the edges of the *a priori* window.

Figure 4(F) shows the trial-by-trial results for the Hackbarth method at a low signal-to-noise ratio. Many of the data points for delay estimates fall on top of the gray bands that mark the edges of the *a priori* window. The analog sum and Saillant methods do not suffer from this problem [Figs. 4(D), (E)]. Although this edge effect was not reported in Hackbarth (1986), she also only plotted results for signal-to-noise ratio values above 20 dB, so we assume she had the same problem in her analysis. Because the Saillant method triggered spikes from *multiple* local peaks above threshold [e.g., Fig. 2(E)], it did not suffer from this edge effect when the low-pass cutoff frequency was >2 kHz.

5. Comparison of the three filterbank methods

Figure 5 shows, for a single 8-kHz low-pass filter condition, the performance of the three filterbank methods on the same plot. From Fig. 5 it was clear that (1) the analog sum method provided the best accuracy (other than the cross-correlation method), and (2) the points where the slope of the curve changed sharply for each method were separated along the horizontal signal-to-noise axis by about 5 dB.

6. Steeper roll-off for the low-pass filter

We also investigated how the severity of low-pass filtering affected the accuracy of the three filterbank delay estimation methods. The second-order Chebyshev low-pass filter had a roll-off of 22 dB/decade. This was rather shallow compared to the 100-dB/decade value estimated for the mammalian smoothing filter (Weiss and Rose, 1988). Therefore, we repeated the simulations using a fifth order Chebyshev low-pass filter that provided attenuation of 62 dB/decade. This more severe low-pass filtering removed most of the ripple riding on each frequency channel's envelope. Consequently, the filterbank performance decreased by 25 dB for all three methods (Fig. 6; Hackbarth method not shown). In addition, the loss of any significant ripple meant that the Saillant method no longer could trigger multiple spikes for the pulse or echo. Consequently, the Saillant and Hackbarth methods converged on using the same single local peak for the pulse and for the echo to generate their spike events, and they yielded similar results.

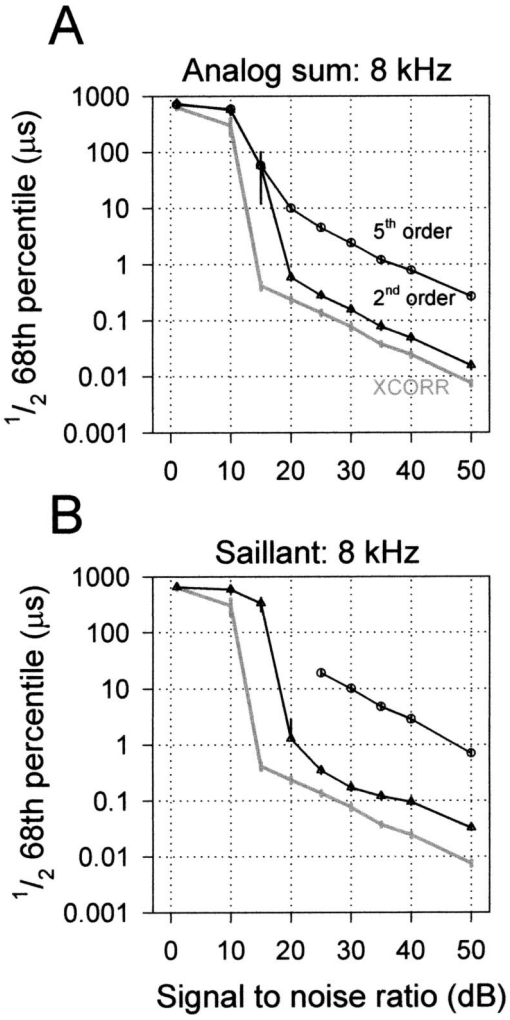


FIG. 6. Increasing the severity of low-pass filtering drastically reduced filterbank accuracy. (A) When the order of the low-pass filter was increased from 2 to 5, the analog sum method's results shifted by about 25 dB. (B) Same as (A), but for the Saillant method.

B. Simulation of the jittering-echo experiment

The performance of the analog sum filterbank model in the jitter experiment is shown in Fig. 7(A). At a signal-to-noise ratio of 36 dB, as in the original behavioral experiment (Simmons *et al.*, 1990), the curves for percent correct show better performance (shift to the left) with increasing cutoff frequency for the low-pass filter. The threshold for jitter detection, Δt_{75}, was taken at the value that resulted in a performance of 75% correct (actual value was identified using cubic spline interpolation).

The thresholds for the cross correlation, analog sum, and Saillant method are shown in Fig. 7(B). The 40-ns threshold, which was measured in the behavioral experiments at a signal-to-noise ratio of 36 dB, is also plotted for comparison. At this signal-to-noise ratio, the best filterbank method (analog sum with a second-order low-pass filter) had a jitter threshold of 82.9 ns. The analog sum method required a rather high SNR, 50 dB, in order to achieve 16.6-ns jitter detection performance based on the two pairs of echoes as simulated stimuli.

IV. CONCLUSIONS

These modeling results show that low-pass filtering strongly affects the accuracy of a filterbank receiver. For bats, this means that achieving 40-ns jitter acuity at a signal-to-noise ratio of ~36 dB requires that cochlear bandpass and low-pass filtering properties adhere to the following two constraints: First, the bandpass filters' integration time should match the sweep rate of the FM emission (discussed previously in Menne, 1988). Second, the effective low-pass filtering that takes place before auditory-nerve spike generation should have, compared to typical mammalian values, a rather high cutoff frequency and shallow slope.

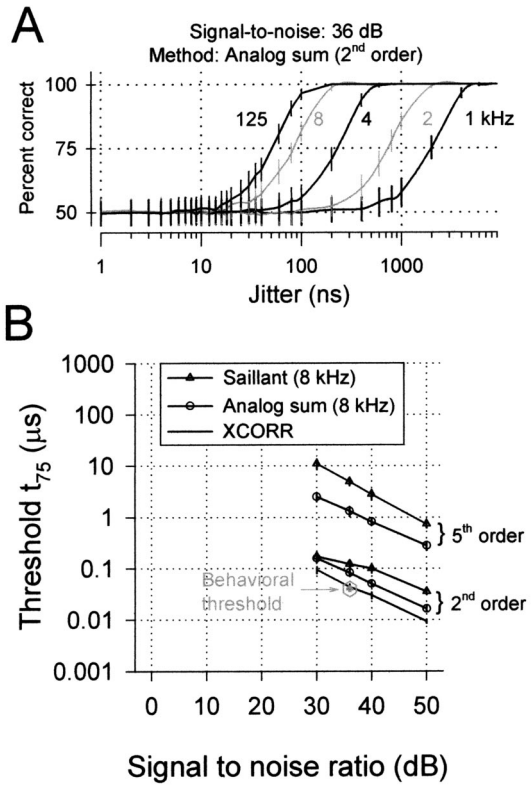


FIG. 7. Jitter threshold for single echo delay. Filterbank delay estimates from the Monte Carlo trials were used to simulate the Simmons *et al.* (1990) single echo jitter experiment (see the text). (A) The performance of the simulated bat in choosing the jittering echo at a signal-to-noise ratio of 36 dB is plotted as percent correct (± st. dev.). The low-pass filter's cutoff frequency is indicated next to each curve. Threshold was taken where the curves crossed the 75%-correct level. (B) The threshold for jitter detection for several different methods. The filterbank models used low-pass filters with an 8-kHz cutoff frequency. The behavioral threshold measured at a signal-to-noise ratio of 36 dB, from Simmons *et al.* (1990), is indicated by the gray symbol.

The simulated jitter experiment assumes the model bat uses the *minimum* number of emissions necessary for the task. In the actual behavioral experiment, the bats certainly took ''multiple looks'' at each side, using anywhere between 6 and 15 or 20 emission per side before making a decision (Simmons *et al.*, 1990). Currently, there is limited knowledge of how bats integrate information across multiple emissions. If the model bat was allowed to average delay estimates across six emissions, accuracy could improve by a factor of 2.45 (in the simulated jitter task, six emissions per side results in three jitter estimates; therefore, the model bat, assuming it had perfect memory storage of those estimates, would see an improvement of $\sqrt{3}$). Applied to the analog sum (second-order) results in Fig. 7 at 36 dB, this would improve the accuracy from 82.9 to 47.9 ns, which is similar to the observed behavioral threshold. At higher signal-to-noise ratios, use of multiple sounds led to comparable improvements in performance.

A. Implementation details for filterbank models

We did not include the effects of internal (receiver) noise upon pulse–echo delay accuracy because modeling the bat's nervous system and cognitive state requires many additional assumptions and parameters that would have to be tested. Internal noise (such as the variability in *when* or *how many* action potentials are generated by auditory neurons, or variability in the memory/decision process) has been considered elsewhere. For example, Suzuki and Suga (1991) assessed the theoretical accuracy for a topographically arranged pool of cortical neurons with variable delay-tuned responses. Their best decoder of cortical activity achieved an echo delay acuity of 800 ns. Another model used to test echo-delay accuracy, developed by Wotton *et al.* (2002), used a population of midbrain, thalamic, and cortical neurons imbued with the latency variability and delay tuning response uncertainty observed in neurophysiological experiments. Similar to Suzuki and Suga's (1991) results, this population model also had a jitter acuity of ~1 μs. Palakal and Wong (1999) also developed a model that used cortical delay-tuned neurons to estimate pulse–echo delay, but did not systematically explore its accuracy beyond reporting that the typical error was about 2%, or ~20 microseconds, of the tested target range. The distribution or variance of these errors was not reported. If possible, future models should incorporate both the effects of internal and external noise on echo delay estimation precision for comparison to the published behavioral data.

This study was motivated in part by an earlier study that documented the low accuracy of a filterbank in the echo-delay estimation task (Hackbarth, 1986). Those earlier results are worse than our filterbank with 1-kHz low-pass filtering (not shown). Several caveats must be mentioned before comparing those data to our simulation results. First, the pulse and echo signals were constructed using a recorded *Eptesicus* broadcast, which had a smaller effective signal bandwidth (it spanned 90–30 kHz) than our test signals. Second, among several differences in the earlier filterbank design, the most significant probably resides in the smoothing procedure. After bandpass filtering, Hackbarth calculated the envelope in a two-step process: she (1) computed the amplitude envelope by taking the absolute value of the signal's Hilbert transform (essentially a full-wave rectification), and (2) low-pass filtered the result at 5 kHz. No further details were provided about this low-pass filter.

The low-pass filter's effect was most deleterious for cutoff frequencies less than the *bandpass* filter bandwidth (i.e., below 4 kHz). When the low-pass filter cutoff was <4 kHz, the low-pass stage not only removes the ac component from the envelope, but more importantly begins to smear the effective integration time established by the bandwidth of the bandpass filter. For a low-pass cutoff above $\geqslant 4$ kHz, some of the ac component passes, and the effective integration time of the whole channel is unchanged, remaining at its minimum value (~250 μs, which originates in the *bandpass* filters' fixed bandwidth of 4 kHz). This surviving ac phase information is crucial for pushing echo-delay accuracy closer to the cross-correlation result (Fig. 4).

B. Low-pass filtering and inner hair cells

The results shown in Fig. 4 reveal the need to obtain constraining empirical data on the bat's inner hair cell membrane filtering properties. In order to move beyond the de-

velopment of the present peripheral auditory filterbank model we need to know both (1) the shape of the effective integration time and (2) the frequency at which ac ripple is no longer detectable in the inner hair cell membrane potential. The available data come from guinea pigs and cats (e.g., Palmer and Russell, 1986), not from bats. Recordings of the inner hair cell receptor potential evoked by pure tones typically show two components: an ac component equal in frequency to the pure tone, and a dc component. The ac component decreases in amplitude with increasing test-tone frequency, while the dc component increases reciprocally. The falloff in the ac amplitude, which for guinea pigs occurs between 0.5–2 kHz (Palmer and Russell, 1986), is due to the inner hair cell's membrane time constant (Russell and Sellick, 1978). Because acquiring data from hair cells is technically difficult, many researchers have addressed this question instead in the auditory nerve. Although the relationship between receptor potential and auditory-nerve spike triggering has not been well studied in a direct manner, for the most part it is reasonable to study phase sensitivity in the auditory nerve because, if it does not exist in the spikes traveling to the cochlear nucleus, then it cannot be detected by the auditory system. In the guinea pig, at least, the decrease in the receptor potential ac component correlates with the decrease in auditory-nerve spike phase locking (Palmer and Russell, 1986). For guinea pig and chinchilla auditory-nerve fibers, phase locking begins to fall off at 0.6 kHz and is negligible by 3.5 kHz (Harrison and Evans, 1979; Palmer and Russell, 1986). Cat auditory-nerve fibers show a gradual falloff that begins around 1–2 kHz and is near zero above 4–5 kHz (Johnson, 1980). Barn owl auditory-nerve fibers show significant phase locking up to 9 kHz (Koppl, 1997). Which of these values is appropriate for FM bats such as *Eptesicus* has yet to be determined. We are not aware of any relevant studies in bats except for one by Suga *et al.* (1971), in which two-tone ''beat'' stimuli evoked phase-locked activity up to 3 kHz in the auditory nerve of *Pteronotus parnelli*.

C. Physiological experiments

Because the neurons of interest for bats have BFs well above 10 kHz (beyond the phase-locking limit seen in birds), phase *locking* to cycles of pure tones cannot be measured. Instead of using long pure tones to measure phase *locking*, a better test for echolocating animals is one of phase *sensitivity*. For bats, important acoustic events are extremely brief FM sounds that typically evoke an average of just 1 spike per stimulus (Pollak *et al.*, 1977; Ferragamo *et al.*, 2002; Sanderson and Simmons, 2000). What matters are the relative latencies of these single spikes across neurons, each of which is evoked by a pulse or echo after a relatively brief period of quiet. Do neurons show latency shifts when the phase of a brief tone burst or FM sweep changes?

Thus far, we have conducted a series of experiments in the brainstem and inferior colliculus of *Eptesicus fuscus* to test this question. Local field potentials do show sensitivity (changes in shape) to changes in pure-tone burst starting phase (Ferragamo *et al.*, 2002). This sensitivity to phase can be observed for pure tones up to 16 kHz. Results from these physiological experiments address findings from experiments in *Eptesicus* that imply (behavioral: Menne *et al.*, 1989; Simmons *et al.*, 1990; Moss and Simmons, 1993) or require (computational: Peremans and Hallam, 1998; Matsuo *et al.*, 2001; Neretti *et al.*, 2003; Saillant *et al.*, 1993) phase sensitivity in the ultrasonic range.

ACKNOWLEDGMENTS

Grants from NSF (BES-9622297) and ONR (N00014-99-1-0350) to J.A.S., an NIH Training grant, and a Burroughs-Wellcome grant to the Brown University Brain Sciences Program supported this work.

Altes, R. A. (**1984**). ''Echolocation as seen from the viewpoint of radar/sonar theory,'' in *Localization and Orientation in Biology and Engineering*, edited by D. Varju and H.-U. Schnitzler (Springer, Berlin), pp. 234–244.

Beedholm, K., and Mohl, B. (**1998**). ''Bat sonar: An alternative interpretation of the 10-ns jitter result,'' J. Comp. Physiol., A **182**, 259–266.

Burdic, W. S. (**1968**). *Radar Signal Analysis* (Prentice-Hall, New York).

Cariani, P. A., and Delgutte, B. (**1996**). ''Neural correlates of the pitch of complex tones. I. Pitch and pitch salience,'' J. Neurophysiol. **76**, 1698–1716.

Dear, S. P., Simmons, J. A., and Fritz, J. (**1993**). ''A possible neuronal basis for representation of acoustic scenes in auditory cortex of the big brown bat,'' Nature (London) **364**, 620–623.

Feng, A. S., Simmons, J. A., and Kick, S. A. (**1978**). ''Echo detection and target ranging neurons in the auditory system of the bat *Eptesicus fuscus*,'' Science **202**, 645–648.

Ferragamo, M. J., Sanderson, M. I., and Simmons, J. A. (**2002**). ''Phase sensitivity of auditory brain-stem responses in echolocating big brown bats,'' J. Acoust. Soc. Am. Suppl. **112**, S-2288.

Grinnell, A. D. (**1995**). ''Hearing in bats: an overview ,'' in *Hearing by Bats*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York), pp. 1–36.

Hackbarth, H. (**1986**). ''Phase evaluation in hypothetical receivers simulating ranging in bats,'' Biol. Cybern. **54**, 281–287.

Harrison, R. V., and Evans, E. F. (**1979**). ''Some aspects of temporal coding by single cochlear fibers from regions of cochlear hair cell degeneration in the guinea pig,'' Arch. Oto-Rhino-Laryngol. **224**, 71–78.

Johnson, D. J. (**1980**). ''The relationship between spike rate and synchrony in responses of auditory nerves to single tones,'' J. Acoust. Soc. Am. **68**, 1115–1122.

Koay, G., Heffner, H. E., and Heffner, R. S. (**1997**). ''Audiogram of the big brown bat (*Eptesicus fuscus*),'' Hear. Res. **105**, 202–210.

Koppl, C. (**1997**). ''Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, *Tyto alba*,'' J. Neurosci. **17**, 3312–3321.

Matsuo, I., Tani, J., and Yano, M. (**2001**). ''A model of echolocation of multiple targets in 3D space from a single emission,'' J. Acoust. Soc. Am. **110**, 607–624.

Menne, D. (**1988**). ''A matched filter bank for time delay estimation in bats,'' in *Animal sonar: Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum Press, New York), pp. 835–842.

Menne, D., Kaipf, I., Wagner, I., Ostwald, J., and Schnitzler, H. U. (**1989**). ''Range estimation by echolocation in the bat *Eptesicus fuscus*: Trading of phase versus time cues,'' J. Acoust. Soc. Am. **85**, 2642–2650.

Menne, D., and Hackbarth, H. (**1986**). ''Accuracy of distance measurement in the bat *Eptesicus fuscus*: Theoretical aspects and computer simulations,'' J. Acoust. Soc. Am. **79**, 386–397.

Moss, C. F., and Schnitzler, H.-U. (**1989**). ''Accuracy of target ranging in echolocating bats: Acoustic information processing,'' J. Comp. Physiol. A **165**, 383–393.

Moss, C. F., and Schnitzler, H.-U. (**1995**). ''Behavioral studies of auditory information processing,'' in *Hearing by Bats*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York), pp. 87–145.

Moss, C. F., and Simmons, J. A. (**1993**). ''Acoustic image representation of a point target in the bat *Eptesicus fuscus*: Evidence for sensitivity to echo phase in bat sonar,'' J. Acoust. Soc. Am. **93**, 1553–1562.

Neretti, N., Sanderson, M. I., Intrator, N., and Simmons, J. A. (**2003**). ''Time-frequency computational model for echo-delay resolution in sonar images of the big brown bat, *Eptesicus fuscus*,'' J. Acoust. Soc. Am. **113**, 2137–2145.
O'Neill, W. E. (**1995**). ''The bat auditory cortex,'' in *Hearing by Bats*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York), pp. 416–480.
O'Neill, W. E., and Suga, N. (**1982**). ''Encoding of target range and its representation in the auditory cortex of the mustached bat,'' J. Neurosci. **2**, 17–31.
Palmer, A. R., and Russell, I. J. (**1986**). ''Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair cells,'' Hear. Res. **24**, 1–15.
Palakal, M. J., and Wong, D. (**1999**). ''Cortical representation of spatiotemporal pattern of firing evoked by echolocation signals: Population encoding of target features in real time,'' J. Acoust. Soc. Am. **106**, 479–490.
Peremans, H., and Hallam, J. (**1998**). ''The spectrogram correlation and transformation receiver revisited,'' J. Acoust. Soc. Am. **104**, 1101–1110.
Pollak, G. D., Marsh, D. S., Bodenhamer, R., and Souther, A. (**1977**). ''Characteristics of phasic-on neurons in the inferior colliculus of anaesthetized bats with observations related to mechanisms for echoranging,'' J. Neurophysiol. **40**, 926–941.
Pollak, G. D. (**1993**). ''Some comments on the proposed perception of phase and nanosecond time disparities by echolocating bats,'' J. Comp. Physiol. A **172**, 523–531.
Russell, I. J., and Sellick, P. M. (**1978**). ''Intracellular studies of hair cells in the mammalian cochlea,'' J. Physiol. (London) **284**, 261–290.
Saillant, P. A., Simmons, J. A., Dear, S. P., and McMullen, T. A. (**1993**). ''A computational model of echo processing and acoustic imaging in frequency-modulated echolocating bats: The spectrogram correlation and transformation receiver,'' J. Acoust. Soc. Am. **94**, 2691–2712.
Sanderson, M. I., and Simmons, J. A. (**2000**). ''Neural responses to overlapping FM sounds in the inferior colliculus of echolocating bats,'' J. Neurophysiol. **83**, 1840–1855.
Schnitzler, H. U., and Henson, O. W. (**1980**). ''Performance of airborne animal sonar systems: Michrochiroptera,'' in *Animal Sonar Systems*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 109–181.
Schnitzler, H.-U., Menne, D., and Hackbarth, H. (**1985**). ''Range determination by measuring time delays in echolocating bats,'' *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer, Berlin), pp. 180–204.
Simmons, J. A. (**1973**). ''Resolution of target range by echolocating bats,'' J. Acoust. Soc. Am. **54**, 157–173.
Simmons, J. A. (**1980**). ''The processing of sonar echoes by bats,'' in *Animal Sonar Systems*, edited by R.-G. Busnel and J. F. Fish (Plenum, New York), pp. 695–714.
Simmons, J. A. (**1993**). ''Evidence for perception of fine echo delay and phase by the FM bat, *Eptesicus fuscus*,'' J. Comp. Physiol., A **172**, 533–547.
Simmons, J. A., Ferragamo, M., Moss, C. F., Stevenson, S. B., and Altes, R. A. (**1990**). ''Discrimination of jittered sonar echoes by the echolocating bat, *Eptesicus fuscus*: The shape of target images in echolocation,'' J. Comp. Physiol., A **167**, 589–616.
Simmons, J. A., and Grinnell, A. D. (**1988**). ''The performance of echolocation: Acousitc images perceived by echolocating bats,'' in *Animal Sonar: Processes and Performance*, edited by P. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 353–385.
Suga, N. (**1971**). ''Responses of inferior collicular neurones of bats to tone bursts with different rise times,'' J. Physiol. (London) **217**, 159–177.
Sullivan, W. E. (**1982**). ''Neural representation of target distance in auditory cortex of the echolocating bat *Myotis lucifugus*,'' J. Neurophysiol. **48**, 1011–1032.
Suzuki, M., and Suga, N. (**1991**). ''Acuity in ranging based upon neural responses in the FM–FM area of the mustached bat,'' Program No. 181.8, Soc. Neuroscience Abstract.
Weiss, T. F., and Rose, C. (**1988**). ''A comparison of synchronization filters in different auditory receptor organs,'' Hear. Res. **33**, 175–188.
Wotton, J., Sanderson, M. I., and Ferragamo, M. J. (**2002**). ''Simulating temporal 'jitter' discriminations using three populations layers of bat auditory neurons,'' Program No. 762.8, Soc. Neuroscience Abstract.

Auditory masking in three pinnipeds: Aerial critical ratios and direct critical bandwidth measurements

Brandon L. Southall,[a] Ronald J. Schusterman,[b] and David Kastak
Long Marine Laboratory, University of California at Santa Cruz, 100 Shaffer Road, Santa Cruz, California 95060

(Received 9 March 2003; accepted for publication 12 May 2003)

This study expands the limited understanding of pinniped aerial auditory masking and includes measurements at some of the relatively low frequencies predominant in many pinniped vocalizations. Behavioral techniques were used to obtain aerial critical ratios (CRs) within a hemianechoic chamber for a northern elephant seal (*Mirounga angustirostris*), a harbor seal (*Phoca vitulina*), and a California sea lion (*Zalophus californianus*). Simultaneous, octave-band noise maskers centered at seven test frequencies (0.2–8.0 kHz) were used to determine aerial CRs. Narrower and variable bandwidth masking noise was also used in order to obtain direct critical bandwidths (CBWs). The aerial CRs are very similar in magnitude and in frequency-specific differences (increasing gradually with test frequency) to underwater CRs for these subjects, demonstrating that pinniped cochlear processes are similar both in air and water. While, like most mammals, these pinniped subjects apparently lack specialization for enhanced detection of specific frequencies over masking noise, they consistently detect signals across a wide range of frequencies at relatively low signal-to-noise ratios. Direct CBWs are 3.2 to 14.2 times wider than estimated based on aerial CRs. The combined masking data are significant in terms of assessing aerial anthropogenic noise impacts, effective aerial communicative ranges, and amphibious aspects of pinniped cochlear mechanics. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1587733]

PACS numbers: 43.80.Lb, 43.80.Nd [WA]

I. INTRODUCTION

A considerable amount is known about certain aspects of acoustic communication and auditory processing in pinnipeds [seals, sea lions, and walruses] (see Ketten and Watzok, 1999). Basic research in pinniped bioacoustics is frequently motivated by the importance of vocal signaling. Of additional interest is the fact that most of these amphibious marine mammals produce and detect both airborne and underwater sounds in environments that favor the use of acoustic communicative signals (Kastak and Schusterman, 1998; Tyack, 1998; Schusterman *et al.*, 2000). Potential interference with these processes by increasing anthropogenic noise levels in some marine environments (see Andrew *et al.*, 2002) is providing impetus for expanding the limited knowledge of the effects of noise on pinniped hearing (Richardson *et al.*, 1995; National Research Council, 2000). Significant interest has focused on the harmful effects of underwater noise (e.g., noise-induced temporary and permanent threshold shifts). Several recent studies have provided data bearing on these issues (Kastak *et al.*, 1999; Schlundt *et al.*, 2000; Finneran *et al.*, 2002). However, we still lack critical data on the effects of airborne sounds, including the potential for anthropogenic noise to mask the relatively low (<2 kHz) frequencies typically contained in many pinniped communicative signals (Richardson *et al.*, 1995; Southall *et al.*, 2000). Therefore, this study focuses on the effects of airborne noise on aerial hearing for a wide range of frequencies, including low frequencies.

An observation from studies involving a variety of mammals is that tonal signals are almost exclusively masked by noise in a narrow band of similar frequencies (Wegel and Lane, 1924; Fletcher, 1940; Scharf, 1970), which Fletcher (1940) refers to as the critical band and its bandwidth as the critical bandwidth (CBW). Based largely on the results of auditory masking studies, mammalian auditory systems are typically viewed as segregating acoustic signals into their constituent frequencies in a manner analogous to the operation of overlapping bandpass filters. Auditory masking is quantified as the critical ratio (CR), which is the difference (in dB) between the masked hearing threshold and the masking noise spectral power density level (Fletcher, 1940; Scharf, 1970). A low CR at a particular frequency indicates relatively efficient extraction of signals from noise compared to a higher CR and, theoretically, a relatively narrower critical CBW. According to Fletcher (1940), the power of a signal at the masked hearing threshold is equal to the total noise power within the critical band. For a particular frequency the CBW may be estimated indirectly as $10^{(CR/10)}$, a method referred to as the critical ratio equal power (CR/EqP) method (as in Richardson *et al.*, 1995). However, studies using a variety of techniques to directly measure auditory filter parameters in human subjects have demonstrated that the CR/EqP method may be of limited accuracy in estimating CBWs (Egan and Hake, 1950; Schafer *et al.*, 1950; Hamilton, 1957; Zwicker *et al.*, 1957; Patterson, 1976).

There are many similarities in mammalian auditory

[a]Current address: National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Acoustics Program, Silver Spring, MD 20910.
[b]Electronic mail: rjschust@cats.ucsc.edu

masking because cochlear mechanics are similar among nonspecialized species (Echteler *et al.*, 1994). There is generally a linear relationship between the masking noise level and effective masking, so that CRs at each frequency are independent of masker level in most conditions (Hawkins and Stevens, 1950; Watson, 1963; Gourevitch, 1965; Johnson, 1968; Ehret, 1976). Additionally, relative auditory filter widths are an approximately constant percentage of center frequency (constant *Q*) but absolute CBWs increase with frequency, resulting in a gradual increase in CRs with frequency (~3 dB/octave).

While these trends are consistent among most mammals, and the relative differences in CR magnitudes are relatively small (see Fay, 1988), certain taxa tend to have somewhat lower CRs on average. For instance, humans, nonhuman primates, and most marine mammals generally have relatively lower CRs than many terrestrial vertebrates tested (Southall *et al.*, 2000). There are also species differences in the relationship between indirectly estimated and directly measured CBWs. In humans, indirect estimates are approximately 2.5 times smaller than directly measured CBWs, a difference that is fairly consistent for many frequencies (see Scharf, 1970). However, studies of nonhuman mammals indicate that this relationship is not found in all species (Seaton and Trahiotis, 1975; Ehret, 1976; Au and Moore, 1990; Turnbull and Terhune, 1990; Niemiec *et al.*, 1992).

Aerial or underwater CRs for individuals from six of the 34 pinniped species have been obtained experimentally (Terhune and Ronald, 1971, 1975; Renouf, 1980; Moore and Schusterman, 1987; Turnbull and Terhune, 1990; Terhune, 1991; Southall *et al.*, 2000). Pinniped CRs are generally low when compared to most other mammals, indicating relatively good hearing in noise. Aerial and underwater pinniped CRs are often directly compared despite the fact that aerial and underwater absolute hearing generally differs. Such comparisons are based on the notion that CRs are related to cochlear processes and that both signals and masking noise should be similarly affected by outer- and middle-ear filtering effects (Renouf, 1980; Turnbull and Terhune, 1990). Indeed, this hypothesis has been confirmed by data for one individual harbor seal, suggesting that there is no difference between pinniped CRs obtained in air and in water (Turnbull and Terhune, 1990). Moreover, these investigators found that critical bandwidths measured directly for the same individual harbor seal were consistently narrower than predicted with the CR/EqP method, became narrower with increasing frequency, and were similar in air and water.

The purpose of the present study was to provide in-air CRs and direct CBW measurements for individuals from three pinniped species: northern elephant seal (*Mirounga angustirostris*), harbor seal (*Phoca vitulina*), and California sea lion (*Zalophus californianus*). Aerial CRs are compared with underwater values for the same individuals obtained in a previous study (Southall *et al.*, 2000). Indirect filter bandwidth estimates are compared with direct CBWs.

II. METHODS

A. Subjects and testing enclosure

The subjects of the aerial auditory masking experiments were an 8-year-old female northern elephant seal (Burnyce), a 13-year-old male harbor seal (Sprouts), and a 16-year-old female California sea lion (Rio). All subjects had many years of experience in audiometric tasks prior to this study (Kastak and Schusterman, 1998; Kastak *et al.*, 1999; Schusterman, unpublished data). None of the subjects had been administered aminogylcoside antibiotics and no permanent hearing losses were known to have occurred. The elephant seal developed a chronic otitis externa in one ear prior to relocation at Long Marine Laboratory. This condition, while intermittently recurrent, is not thought to significantly impact absolute aerial or underwater hearing (see Kastak and Schusterman, 1998).

The animals were housed at Long Marine Laboratory in Santa Cruz, California in seawater-filled pools and adjacent haul-out space. Subjects received 25%–75% of their daily allotment of fish during one to three experimental sessions, 5 to 7 mornings per week between October 2001 and January 2002. The experiments were conducted under National Marine Fisheries Service (NMFS) permit 259-1481-00 and were approved by the University of California, Santa Cruz Chancellor's Animal Research Committee.

A hemianechoic chamber (Eckel Industries) mounted on a semi-isolated cement foundation served as the testing enclosure. The chamber consisted of a 3.3×2.3×2.2-m test room and an adjacent 1.3×3.3-m experimental control room. The outside walls of the entire chamber were double-paneled, 8-gauge stainless steel, providing 40–60-dB of noise attenuation. The inside walls and ceiling of the test room were lined with fiberglass-filled, aluminum-covered anechoic wedges that limited acoustic reflection within the testing room. The cement floor of the test room was covered with several 2.6-cm-thick neoprene mats. The inside walls of the control room were lined with sound-absorptive foam panels. Equipment cable ducts between the test and control rooms were lined with sound-absorptive foam.

A floor-mounted convex section of polyvinyl chloride (PVC) served as a chin station for experimental sessions. An acoustically sealed, sliding PVC feeding tube allowed fish reinforcement to be delivered to the subjects from the control room during sessions. Animal responses were observed via a closed-circuit video system. A square plastic response paddle was located 0.3 m in front of and to the left of the chin station. The chin station was positioned in the same location for each individual, approximately 1 m in front of one of several projecting speakers, to ensure consistent placement of subjects within and between sessions.

B. Test stimuli and masking noise

Test stimuli were pure tones at seven frequencies between 0.2 and 8.0 kHz with either 20- or 40-ms rise/fall times and 500-ms overall duration. Two hardware and software configurations were used to generate these signals. The initial system was identical to that previously used for underwater auditory masking experiments (Southall *et al.*, 2000).

For some of the test conditions, including the entire band-narrowing experiment, signals were generated using custom-built software (LABVIEW) and a 6070E multifunction board mounted in a PXI-1010 chassis (National Instruments). For both configurations, signals were projected using either a JBL 2245H (0.2 kHz) JBL 2123H (0.5, 0.8, and 1.2 kHz), or JBL 2123H/J (2.5, 4.0, and 8.0 kHz) speaker.

Variable bandwidth masking noise was generated using COOL EDIT software (Syntrillium). Identical 6-s noise intervals were produced for each experimental trial using the sound card of a personal computer (sampling rate 22 050 Hz, 16-bit resolution, FFT size 512). The sound card output was fed through a Krohn-Hite 3550 bandpass filter, a Hewlett Packard stepwise attenuator, and into another input channel of the power amplifier, which drove the projecting transducer. Test signals and masking noise were projected from the same source to avoid reduced masking resulting from angular separation between signal and masker sources (Turnbull, 1994). Masking noise was presented only for the duration of each experimental trial was to prevent any confounding effects resulting from loudness adaptation (see Southall *et al.*, 2000).

At each frequency a single masking noise level was used, based on subject's minimum audible field hearing threshold measured in the anechoic chamber (Schusterman *et al.*, unpublished data). It was not possible to use the same relative noise levels for every frequency, due to limitations of the equipment and the fact that absolute aerial thresholds often varied greatly between subjects. Masking noise at each frequency and noise bandwidth had an average spectral density level [dB *re*: (20 μPa)^2/Hz] approximately 5–20-dB above each subject's hearing threshold. Masking noise was digitally filtered so that received masking noise spectral density levels were relatively flat ($\pm$3 dB) across the entire masker band.

Signal sound-pressure levels (dB *re*: 20 μPa) and masking noise spectral density levels [dB *re*: (20 μPa)^2/Hz] were calibrated before and after each session. An Etymotic ERC-7 clinical probe microphone with an internal reference calibration, a Tectronix TDS 420A digital oscilloscope, and SPECTRA PLUS software were used to analyze signal and noise characteristics. In addition to calibrating signal and masking noise levels at a central position corresponding to the center of the subject's head during testing, signal and noise fields were mapped at multiple nearby positions (as in Moore and Schusterman, 1987; Southall *et al.*, 2000). Mapping was conducted at 27 positions within a 10$\times$10$\times$10-cm region around the chin station by: (1) measuring received signal and masker levels; (2) analyzing the distribution of noise energy within masker bands; and (3) observing how these parameters varied within or between masker intervals. Signal and masking noise band average spectral density levels at each position were maintained within 3-dB of the average levels measured at the central calibration position both within and between multiple masker intervals.

C. Procedure

The psychophysical procedures employed were similar to those used by Kastak and Schusterman (1998) and Southall *et al.* (2000). The use of experienced test subjects resulted in little additional training for the aerial masking experiments. Once subjects stationed appropriately, masking noise was engaged and a trial light placed near the projecting speaker was illuminated for the 6-s duration of each experimental trial. Masking noise alone was presented on either 25% or 50% of the trials (catch trials). Signals in addition to masking noise were projected on the remaining trials (signal trials). Either a 75:25 or 50:50 signal-to-catch trial ratio was used in to maintain appropriate subject response criteria. The order of signal and catch trials varied between sessions based on computer-generated 60-trial pseudorandom sequences. A go/no-go psychophysical response paradigm was used. Subjects reported detecting a signal by leaving the chin station and pressing the response paddle with their muzzle; they reported not detecting a signal by withholding this response. Fish reinforcement was delivered following correct responses [responding on a signal trial (hit) or not responding on a catch trial (correct rejection)]. Subjects restationed without reinforcement following incorrect responses [not responding on a signal trial (miss) or responding on a catch trial (false alarm)].

Each session consisted of an 8–20-trial warm-up phase in which signals were clearly audible, a 20–40-trial threshold phase, and a terminal 4–6-trial cool-down phase with clearly audible signals. The warm-up and cool-down phases were used to assess subject motivation and to maintain stimulus control over behavioral responses.

Two psychophysical techniques were used in determining masked hearing thresholds. First, three to eight sessions using the staircase method (Cornsweet, 1962) were conducted to obtain an initial estimate of the masked hearing threshold at each frequency. Signal levels were initially attenuated by 4-dB following each hit and were then adjusted in 2-dB steps in the threshold phase (increased following a miss, decreased following a hit), which consisted of nine reversals between sequences of hits and misses. Subsequently, five to seven discrete signal levels separated by 2-dB were selected, based on the staircase data, and randomly presented during three to five method-of-constant-stimuli sessions (Stebbins, 1970). Masked hearing thresholds were determined by pooling results from all method-of-constant-stimuli sessions and calculating the 50%-correct detection point using probit analysis (Finney, 1971). When a threshold's 95%-confidence limits fell within 3-dB of the calculated masked hearing threshold, testing at a particular frequency was ended.

1. Aerial critical ratio procedure (octave-band noise)

Aerial masked hearing thresholds were determined in the presence of octave band noise (OBN) maskers centered at test frequencies of 0.2, 0.5, 0.8, 1.2, 2.5, 4.0, and 8.0 kHz. The sequence of testing was randomized with respect to test frequency, but the same order was used for each test subject.

Critical ratios were calculated as the difference (in dB) between the masked hearing threshold sound-pressure level and the masking noise sound-pressure level at the center frequency of the masking band, corresponding to the measured masker spectral density level [dB *re*: (20 μPa)^2/Hz]

TABLE I. Aerial auditory masking data obtained using octave-band noise (70.7% of center frequency) for individual subjects from three pinniped species tested in a hemianechoic chamber. Data were obtained using the psychophysical method of constant stimuli.

Species	Frequency (kHz)	Masker level (dB *re*: 20 μPa^2/Hz)	Masked threshold (dB *re*: 20 μPa) & (std dev)	False-alarm rate (%)	Critical ratio (dB)
N. elephant seal	0.2	45.0	57.1 (1.0)	5.6	12
Harbor seal	0.2	41.0	53.3 (1.1)	12.8	12
California sea lion	0.2	…	…	…	…
N. elephant seal	0.5	31.0	44.4 (1.2)	5.9	13
Harbor seal	0.5	31.0	49.2 (0.8)	21.0	18
California sea lion	0.5	38.0	58.3 (0.8)	11.9	20
N. elephant seal	0.8	42.7	56.6 (1.3)	15.0	14
Harbor seal	0.8	33.8	49.6 (0.9)	13.3	16
California sea lion	0.8	45.4	62.6 (0.9)	9.8	17
N. elephant seal	1.2	37.5	55.4 (0.9)	18.2	18
Harbor seal	1.2	28.0	49.9 (0.9)	11.8	22
California sea lion	1.2	33.0	52.7 (1.1)	19.2	20
N. elephant seal	2.5	35.0	54.7 (0.9)	19.4	20
Harbor seal	2.5	20.0	40.4 (1.0)	17.6	20
California sea lion	2.5	27.0	45.7 (1.9)	8.3	19
N. elephant seal	4.0	40.0	62.1 (0.8)	18.8	22
Harbor seal	4.0	17.0	38.8 (0.8)	17.8	22
California sea lion	4.0	35.0	60.1 (1.3)	10.5	25
N. elephant seal	8.0	40.0	63.5 (0.8)	12.5	24
Harbor seal	8.0	25.0	46.7 (1.3)	6.3	22
California sea lion	8.0	25.0	51.2 (1.7)	11.5	26

(Fletcher, 1940; Scharf, 1970). The CRs obtained using octave band masking noise were used to indirectly estimate CBWs and additionally served as the initial broadband measurement for direct CBW measurements (described below).

2. Direct aerial critical bandwidth procedure (variable bandwidth noise)

Following the OBN masking experiments, two or three additional method-of-constant-stimuli sessions were conducted for five relatively narrower masker bandwidths. This procedure was conducted at the three highest frequencies (2.5, 4.0, and 8.0 kHz) to minimize the potential for audible interactions between tonal signals and narrow-band noise maskers (Bos and de Boer, 1966). Critical ratios were calculated in the same manner described above for masking noise with bandwidths of 3%, 6%, 12%, 18%, and 24% of each test frequency. All sessions for each masker bandwidth were conducted sequentially, but the testing order for each relative masker bandwidth was varied between frequencies. Critical bandwidths were estimated at each frequency based on CRs for different masker bandwidths. Specifically, CBWs as a percentage of center frequency were estimated as the intersection of the least-squares estimates functions for the fairly constant CRs (within 1-dB of CRs determined using OBN) at wide noise bandwidths and for the relatively lower CRs (more than 1-dB below the OBN value) at narrower masker bandwidths (as in Seaton and Trahiotis, 1975; Au and Moore, 1990; Niemiec *et al.*, 1992).

III. RESULTS

A. Aerial critical ratios

For each of the three pinniped species tested, the aerial masked hearing thresholds as well as the CRs obtained using OBN for each test frequency are shown in Table I. Along with these values, the percentage of false alarms pooled for all method-of-constant-stimuli sessions is given for each animal. The aerial CRs for these three pinnipeds are quite similar to those determined by Southall *et al.* (2000) at overlapping frequencies under water (Fig. 1). For each subject, CRs generally increase gradually with frequency.

For the northern elephant seal, aerial CRs range from 12-dB at 0.2 kHz to 24-dB at 8.0 kHz and are not significantly different than underwater CRs at overlapping test frequencies. For the harbor seal, aerial CRs range from 12-dB at 0.2 kHz to 22-dB at 4.0 kHz and are not significantly different than underwater CRs at overlapping test frequencies. For the California sea lion, aerial CRs range from 17-dB at 0.8 kHz to 26-dB at 8.0 kHz and are not significantly different than underwater CRs at overlapping test frequencies.

False-alarm rates reflect subject response bias. High false-alarm rates indicate a relatively liberal response criterion, and low false-alarm rates indicate a conservative response criterion. We adjusted the relative percentage of signal and catch trials as necessary to ensure fairly consistent false alarm rates (between approximately 5%–20%) across test frequencies (Table I). Subjects who adopted response criteria that were either too liberal (>25% false-alarm rates on more than three consecutive sessions) or too conservative (0% false-alarm rates on more than three consecutive sessions) generally responded rapidly and appropriately to re-

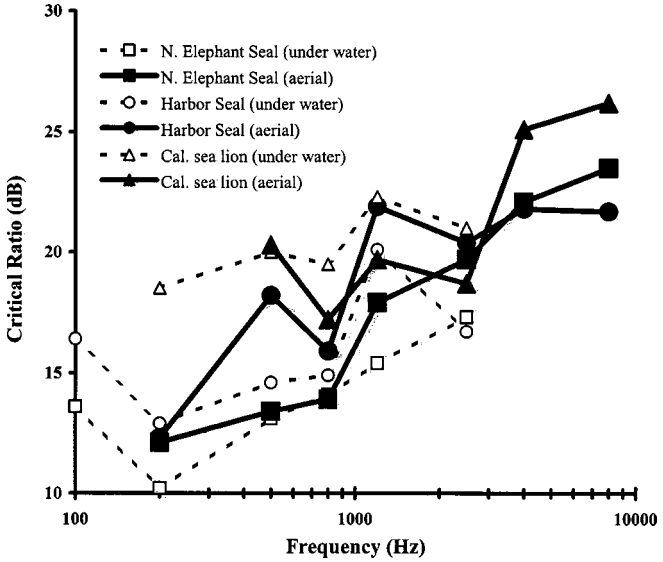


FIG. 1. Critical ratios versus frequency for aerial (this study) and underwater auditory masking (Southall *et al.*, 2000) for a northern elephant seal, a harbor seal, and a California sea lion.

spective decreases or increases in the relative signal-to-catch trial ratio.

B. Direct aerial critical bandwidths

Critical ratios at relatively wide masking noise bandwidths are very similar to those obtained using OBN, while lower CRs occur for relatively narrow masking noise bandwidths. At each test frequency, directly measured aerial CBWs are much wider than indirect estimates based on the aerial CRs obtained with OBN (Fig. 2). However, there are no consistent relationships across the three frequencies tested, between direct and indirect CBWs for any of the test subjects. For the northern elephant seal, direct aerial CBWs are 6.6, 3.9, and 4.6 times wider than the indirect estimates

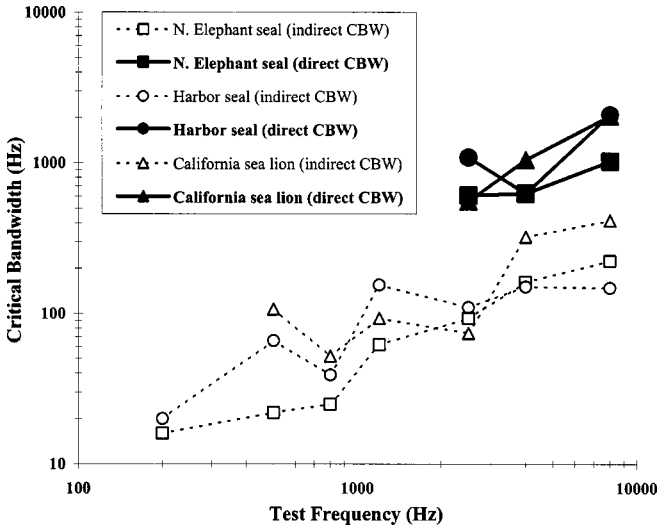


FIG. 2. Critical bandwidths (CBW) vs test frequency obtained in air (this study) and under water (Southall *et al.*, 2000) based on indirect estimates (CR/EqP method) and direct measurements obtained with a band-narrowing technique for a northern elephant seal, a harbor seal, and a California sea lion.

based on aerial CRs at 2.5, 4.0, and 8.0 kHz, respectively. The direct CBWs for this individual are 24.4%, 15.8%, and 12.9% of these test frequencies. For the harbor seal, direct aerial CBWs are 9.9, 4.2, and 14.2 times wider than the indirect estimates and equal 43.6%, 16.0%, and 26.3% of the respective test frequencies. For the California sea lion, direct aerial CBWs are 7.6, 3.2, and 4.9 times wider than the indirect estimates and equal 22.4%, 26.3%, and 25.8% of the respective test frequencies.

As in the CR measurements using OBN, pooled false-alarm rates in the critical bandwidth study were generally between 5%–20% across test frequencies. Interestingly though, for each subject, false alarms occurred in some cases during the first several catch trials using the two narrowest bandwidth maskers. This was generally not observed during the initial presentations of relatively wider masker bands. False alarms on the first few catch trials with the narrow-band maskers probably occurred because narrow-band noise is qualitatively more similar to test stimuli than wide bands of noise. After a few trials this behavior generally ceased, as subjects rapidly learned the distinctions between test signals and narrow-band noise.

IV. DISCUSSION

Aerial CRs obtained in the present study are comparable to underwater values for the same individuals (Southall *et al.*, 2000). This replication of Turnbull and Terhune's (1990) findings provides additional evidence that the masking effects of noise are similar in air and water. Because aerial and underwater CRs are similar for these particular species, which have quite different amphibious hearing characteristics and life histories (see Kastak and Schusterman, 1998), this trend likely applies for most, if not all, other pinniped species. Additionally, these results suggest that pinniped cochlear mechanisms are similar in air and water. This finding suggests that direct comparisons of aerial and underwater masked hearing capabilities between aquatic, amphibious, and terrestrial species may be appropriate (as in Southall *et al.*, 2000). However, comparative aerial and underwater masking data for a greater number of species with variable life history characteristics are needed to more fully address the validity of such comparisons.

While the differences are fairly small, the aerial CRs measured in this study are relatively low in magnitude compared to most terrestrial mammals (see Fay, 1988). This trend is also observed in most other pinnipeds [except the ringed seals tested thus far (Terhune and Ronald, 1975)] and the odontocete cetaceans tested thus far (see Richardson *et al.*, 1995; Southall *et al.*, 2000). It is reasonable to speculate that acoustic signal production and reception in typically noisy marine environments has led to selection for enhanced ability to detect signals in noise (Schusterman *et al.*, 2000). However, additional data are needed, including further research on ringed seals, to comprehensively assess this hypothesis.

Also evident in this study and consistent with previous results is that CRs increase monotonically with frequency. Thus, pinnipeds do not appear to be specialized to detect specific tonal signals, including those predominant in most

pinniped vocalizations. The aerial CRs increase at a similar rate as in most other mammals (see Fay, 1988), supporting the notion that pinnipeds are hearing generalists with respect to frequency processing (Southall *et al.*, 2000).

The presence of critical bands in the pinniped subjects of this study provides further indication that pinniped cochlear processes are similar to those occurring in other generalist mammals. Direct CBWs obtained in this study were much wider than indirect estimates. This finding is similar to data obtained using the band-narrowing technique in humans (e.g., Hamilton, 1957) and a bottlenose dolphin (Au and Moore, 1990), but opposite the relationships determined for chinchillas using the band-narrowing method (Niemiec *et al.*, 1992) and a harbor seal using tonal maskers (Turnbull and Terhune, 1990). As observed in other marine mammals for which direct CBWs have been determined (Au and Moore, 1990; Turnbull and Terhune, 1990), there were not consistent quantitative relationships across test frequencies between direct and indirect CBWs for the pinniped subjects in this study. However, direct CBWs did not become relatively narrower with increasing frequency. The direct CBWs were roughly similar (~13%–44% of test frequencies) to those determined for humans (~15%–22%) using band narrowing (Hamilton, 1957; Greenwood, 1961; van den Brink, 1964), corresponding to approximately 1/3-octave bands. However, the indirect aerial CBW estimates were generally much narrower for these pinnipeds (~3%–9% of test frequencies), because of the relatively low CRs from which they are calculated, corresponding to approximately 1/12-octave bands.

There are several methodological limitations in utilizing the band-narrowing technique for direct CBW measurements. Since masking noise is concentrated in a relatively narrow bandwidth, subjects may detect unintended harmonics of tonal signals occurring outside the noise band, called off-frequency listening. Also, when the masking band becomes particularly narrow, where the most important measurements are made, there can be perceptible interactions between signal and masker as a result of audible amplitude fluctuations occurring in narrow-band noise (Bos and de Boer, 1966). Additionally, CBWs determined using the band-narrowing method depend on the slope of the least-squares estimate through the variable CRs obtained in relatively narrow-band noise, which may be determined by as few as two data points. Despite these limitations, however, Scharf (1970) and Seaton and Trahiotis (1975) asserted that the band-narrowing method provides fairly accurate CBWs for humans, noting the similarity of CBWs determined using band narrowing (Hamilton, 1957; Greenwood, 1961; van den Brink, 1964) with the generally accepted results that Zwicker *et al*. (1957) obtained using the loudness summation technique. Additional methods of directly measuring CBWs (e.g., notched and/or rippled noise maskers) should be used in future studies of pinniped masking in order to more fully understanding auditory filter bandwidths and shapes (as in Halpern and Dallos, 1986; Niemiec *et al*., 1992).

The combined aerial and underwater masking data provide much needed information on the simultaneous effects of noise on hearing for pinnipeds at frequencies contained in many acoustic communicative signals. The masking data will be useful to researchers, regulatory or military agencies, and environmental groups interested in generating quantitative estimates of communication ranges in natural or anthropogenic noise conditions. Further, filter bandwidth measurements, as well as the comparative aerial and underwater CRs, clarify cochlear processes in these amphibious mammals. Finally, the estimated CBWs may be used to define biologically significant filter bandwidths for analyzing communicative signals as well as the effects of natural or anthropogenic noise (Southall *et al.*, in press).

ACKNOWLEDGMENTS

This research was supported by Grant No. N000140210159 from the Office of Naval Research and an Augmentation Award for Science and Engineering Research Training grant from the Department of Defense to R.J.S., as well as grants from the University of California, Santa Cruz Department of Ocean Sciences and Friends of Long Marine Laboratory to B.L.S. The authors acknowledge the assistance of Colleen Reichmuth Kastak, Marla Holt, Kirsten Jensen, and the student volunteers of the Pinniped Research in Cognition and Sensory Systems at Long Marine Laboratory. The authors also appreciate the comments of two anonymous reviewers.

Andrew, R. K., Howe, B. M., Mercer, J. A., and Dzieciuch, M. A. (**2002**). ''Ocean ambient sound: Comparing the 1960s with the 1990s for a receiver off the California coast,'' ARLO **3**, 65–70.

Au, W. W. L., and Moore, P. W. B. (**1990**). ''Critical ratio and critical bandwidth for the Atlantic bottlenose dolphin,'' J. Acoust. Soc. Am. **88**, 1635–1638.

Bos, C. E., and de Boer, E. (**1966**). ''Masking and discrimination,'' J. Acoust. Soc. Am. **39**, 708–715.

Clack, T. D. (**1966**). ''Effect of signal duration on the auditory sensitivity of humans and monkeys (*Macaca mulatta*),'' J. Acoust. Soc. Am. **40**, 1140–1146.

Cornsweet, T. N. (**1962**). ''The staircase method in psychophysics,'' Am. J. Psychol. **75**, 485–491.

Echteler, S. M., Fay, R. R., and Popper, A. N. (**1994**). ''Structure of the mammalian cochlea,'' in *Comparative Hearing: Mammals*, edited by R. R. Fay and A. N. Popper (Springer, New York), pp. 134–171.

Egan, J. P., and Hake, W. (**1950**). ''On the masking pattern of a simple auditory stimulus,'' J. Acoust. Soc. Am. **22**, 622–630.

Ehret, G. (**1976**). ''Critical bands and filter characteristics in the ear of the housemouse (*Mus musculus*),'' Biol. Cybern. **24**, 35–42.

Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay, Winnetka, IL).

Finneran, J. J., Schlundt, C. E., Dear, R., Carder, D. A., and Ridgway, S. H. (**2002**). ''Temporary shift in masked hearing thresholds in odontocetes after exposure to single underwater impulses from a seismic watergun,'' J. Acoust. Soc. Am. **111**, 2929–2940.

Finney, D. J. (**1971**). *Probit Analysis*, 3rd ed. (Cambridge University Press, Cambridge).

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.

Gourevitch, G. G. (**1965**). ''Auditory masking in the rat,'' J. Acoust. Soc. Am. **37**, 439–443.

Greenwood, D. D. (**1961**). ''Auditory masking and the critical band,'' J. Acoust. Soc. Am. **33**, 484–502.

Halpern, D. L., and Dallos, P. (**1986**). ''Auditory filter shapes in the chinchilla,'' J. Acoust. Soc. Am. **80**, 765–775.

Hamilton, P. M. (**1957**). ''Noise masked thresholds as a function of tonal duration and masking noise bandwidth,'' J. Acoust. Soc. Am. **29**, 506–511.

Hawkins, Jr., J. E., and Stevens, S. S. (**1950**). ''The masking of pure tones and of speech by white noise,'' J. Acoust. Soc. Am. **22**, 6–13.

Johnson, C. S. (**1968**). ''Masked tonal thresholds in the bottlenosed porpoise,'' J. Acoust. Soc. Am. **44**, 965–967.
Kastak, D., Schusterman, R. J., Southall, B. L., and Reichmuth, C. J. (**1999**). ''Underwater temporary threshold shift in three species of pinniped,'' J. Acoust. Soc. Am. **106**, 1142–1148.
Kastak, D., and Schusterman, R. J. (**1998**). ''Low-frequency amphibious hearing in pinnipeds: Methods, measurements, noise, and ecology,'' J. Acoust. Soc. Am. **103**, 2216–2228.
Ketten, D., and Wartzok, D. (**1999**). ''Marine mammal sensory systems,'' *The Biology of Marine Mammals*, edited by J. E. Reynolds III and S. A. Rommel (Smithsonian Institution Press, Washington), pp. 117–175.
Moore, P., and Schusterman, R. J. (**1987**). ''Audiometric assessment of northern fur seals (*Callorhinus ursinus*),'' Marine Mammal Sci. **3**, 31–53.
National Research Council. (**2000**). *Marine Mammals and Low-Frequency Sound: Progress since 1994* (National Academy Press, Washington D.C.).
Niemiec, A. J., Yost, W. A., and Shofner, W. P. (**1992**). ''Behavioral measures of frequency selectivity in the chinchilla,'' J. Acoust. Soc. Am. **92**, 2636–2649.
Patterson, R. D. (**1976**). ''Auditory filter shapes derived with noise stimuli,'' J. Acoust. Soc. Am. **59**, 640–654.
Renouf, D. (**1980**). ''Masked hearing thresholds of harbour seals (*Phoca vitulina*) in air,'' J. Aud Res. **20**, 263–269.
Richardson, J. W., Greene, C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, New York).
Schafer, T. H., Gales, R. S., Shewmaker, C. A., and Thompson, P. O. (**1950**). ''The frequency selectivity of the ear as determined by masking experiments,'' J. Acoust. Soc. Am. **22**, 490–496.
Scharf, B. (**1970**). ''Critical bands,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 1, pp. 159–202.
Schlundt, C. E., Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2000**). ''Temporary shift in masked hearing thresholds (MTTS) of bottlenose dolphins, *Tursiops truncatus*, and white whales, *Delphinapterus leucas*, after exposure to intense tones,'' J. Acoust. Soc. Am. **107**, 3496–3508.
Schusterman, R. J. (**1978**). ''Vocal communication in pinnipeds,'' in *Behavior of Captive Wild Animals*, edited by H. Markowitz and V. J. Stevens (Nelson-Hall, Chicago), pp. 247–308.
Schusterman, R. J. (unpublished data).
Schusterman, R. J., Kastak, D., Levenson, D. H., Reichmuth, C. J., and Southall, B. L. (**2000**). ''Why pinnipeds don't echolocate,'' J. Acoust. Soc. Am. **107**, 2256–2264.
Seaton, W. H., and Trahiotis, C. (**1975**). ''Comparison of critical ratios and critical bands in the monaural chinchilla,'' J. Acoust. Soc. Am. **57**, 193–199.
Southall, B. L., Schusterman, R. J., and Kastak, D. **in press**. ''Acoustic communication ranges in northern elephant seals,'' Aquat. Mamm.
Southall, B. L., Schusterman, R. J., and Kastak, D. (**2000**). ''Masking in three pinnipeds: Underwater, low-frequency critical ratios,'' J. Acoust. Soc. Am. **108**, 1322–1326.
Stebbins, W. C. (**1970**). ''Principles of animal psychophysics,'' in *Animal Psychophysics: The Design and Conduct of Sensory Experiments*, edited by W. C. Stebbins (Appleton-Century-Crofts, New York), pp. 1–19.
Terhune, J. M. (**1991**). ''Masked and unmasked pure tone thresholds of a harbor seal listening in air,'' Can. J. Zool. **69**, 2059–2066.
Terhune, J., and Ronald, K. (**1971**). ''The harp seal, *Pagophilus groenlandicus* (Erxleben 1777). The air audiogram,'' Can. J. Zool. **49**, 385–390.
Terhune, J. M., and Ronald, K. (**1975**). ''Masked hearing thresholds of ringed seals,'' J. Acoust. Soc. Am. **58**, 515–516.
Turnbull, S. D. (**1994**). ''Changes in masked thresholds of a harbor seal (*Phoca* vitulina) associated with angular separation of signal and noise sources,'' Can. J. Zool. **72**, 1863–1866.
Turnbull, S. D., and Terhune, J. (**1990**). ''White noise and pure tone masking of pure tone thresholds of a harbor seal listening in air and under water,'' Can. J. Zool. **68**, 2090–2097.
Tyack, P. L. (**1998**). ''Acoustic communication under the sea,'' in *Animal Acoustic Communication*, edited by S. L. Hopp, M. J. Owren, and C. S. Evans (Springer, Berlin), pp. 163–220.
van den Brink, G. (**1964**). ''Detection of tone pulses of various durations in noise of various bandwidth,'' J. Acoust. Soc. Am. **36**, 1206–1211.
Watson, C. S. (**1963**). ''Masking of tones by noise for the cat,'' J. Acoust. Soc. Am. **35**, 167–172.
Wegel, R. L., and Lane, C. E. (**1924**). ''The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear,'' Phys. Rev. **2**, 266–285.
Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). ''Critical bandwidth in loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.

Auditory and behavioral responses of California sea lions (*Zalophus californianus*) to single underwater impulses from an arc-gap transducer[a)]

James J. Finneran
Space and Naval Warfare Systems Center, San Diego, Code 2351, 53560 Hull Street, San Diego, California 92152-5001

Randall Dear
Science Applications International Corporation, Maritime Services Division, 3990 Old Town Avenue, Suite 105A, San Diego, California 92110

Donald A. Carder and Sam H. Ridgway
Space and Naval Warfare Systems Center, San Diego, Code 2351, 53560 Hull Street, San Diego, California 92152-5001

(Received 17 May 2002; revised 19 May 2003; accepted 27 May 2003)

A behavioral response paradigm was used to measure underwater hearing thresholds in two California sea lions (*Zalophus californianus*) before and after exposure to underwater impulses from an arc-gap transducer. Preexposure and postexposure hearing thresholds were compared to determine if the subjects experienced temporary shifts in their masked hearing thresholds (MTTS). Hearing thresholds were measured at 1 and 10 kHz. Exposures consisted of single underwater impulses produced by an arc-gap transducer referred to as a ''pulsed power device'' (PPD). The electrical charge of the PPD was varied from 1.32 to 2.77 kJ; the distance between the subject and the PPD was varied over the range 3.4 to 25 m. No MTTS was observed in either subject at the highest received levels: peak pressures of approximately 6.8 and 14 kPa, rms pressures of approximately 178 and 183 dB *re*: 1 μPa, and total energy fluxes of 161 and 163 dB *re*: 1 μPa^{2}s for the two subjects. Behavioral reactions to the tests were observed in both subjects. These reactions primarily consisted of temporary avoidance of the site where exposure to the PPD impulse had previously occurred. [DOI: 10.1121/1.1598194]

PACS numbers: 43.80.Nd, 43.80.Lb [WA]

I. INTRODUCTION

Pinnipeds are known to negatively impact the commercial passenger fishing vessel (CPFV) industry in southern California through depredation of gear and catch. A number of measures have been used to nonlethally prevent sea lion interaction with CPFVs, including gunshots, small explosive charges (''seal bombs''), and electronic sound generators (e.g., acoustic harassment devices, or AHDs) (Richardson *et al.*, 1995; Reeves *et al.*, 1996; NMFS, 1997). Unfortunately, none of the available nonlethal deterrent methods has been completely effective. Many acoustic deterrents are initially effective but become less so over time as the sea lions become habituated to the sound, learn to ignore or avoid the sound, or actually become less sensitive to the sound (e.g., through permanent hearing damage). In some cases acoustic deterrents may have actually alerted the sea lions to the presence of the CPFV (and its associated fish)—the so called ''dinner bell'' effect (Richardson *et al.*, 1995).

In a variation of the AHD concept, Shaughnessy *et al.* (1981) used an arc-gap transducer to generate underwater shock waves in an effort to deter Cape fur seals (*Arctocephalus pusillus*) from fishing nets. The arc-gap transducer, or sparker, produces an electric spark which vaporizes the water between two submerged electrodes. This causes gas bubbles to form and then quickly collapse, producing an underwater shock wave. At a stored electrical energy setting of 520 J, the device was capable of producing a shock wave with a peak pressure of 132 dB *re*: 1 μPa (~4 Pa) at 1 m. The device appeared effective at close range (2–10 m) but was ineffective at larger distances. In 1997, NOAA Fisheries contracted the Pacific States Marine Fisheries Commission (PSMFC) to develop, build, and test an arc-gap transducer capable of producing larger peak pressures than the device used by Shaughnessy *et al.*, which would presumably give the device a larger effective range. The resulting transducer, known as a pulsed power device (PPD), operates similarly to an arc-gap transducer but is capable of storing up to 3 kJ of electrical energy, compared to the 0.52 kJ of the Shaughnessy *et al.* device.

This paper presents the results of a study designed to investigate the potential auditory effects of exposure to impulses produced by the PPD on California sea lions (*Zalophus californianus*). Marine mammals exposed to sufficiently intense underwater impulses may suffer hearing loss, called a noise-induced (hearing) threshold shift (NITS) or simply a threshold shift (TS). A TS may be a temporary threshold shift (TTS) or a permanent threshold shift (PTS).

[a)]Portions of these data were presented in ''Review of marine mammal temporary threshold shift (TTS) measurements and their application to damage-risk criteria,'' at the 142nd Meeting of the Acoustical Society of America, Fort Lauderdale, FL, December 2001.

TTS is associated with a temporary hearing loss; PTS indicates permanent hearing loss. There have been two studies of TTS in pinnipeds: Kastak and Shusterman (1996) reported TTS in a harbor seal (*Phoca vitulina*) exposed to airborne noise from nearby construction. Kastak *et al.* (1999) reported TTS in a California sea lion, harbor seal, and Northern elephant seal (*Mirounga angustirostris*) exposed to underwater octave band noise.

The auditory effects of the PPD were assessed by using a behavioral response paradigm to measure underwater hearing thresholds in trained sea lions before and after exposure to single underwater impulses produced by the PPD. The preexposure and postexposure hearing thresholds were then compared to see if a TTS had occurred. Tests were conducted at several combinations of PPD electrical charge and distance from the sea lion.

II. METHODS

A. Experimental subjects

Experimental subjects consisted of two male California sea lions: NRT (age 19 years, approximate weight 120 kg) and LIB (23 years, 150 kg). Neither subject had participated in any previous hearing studies. Subjects were housed in a floating netted enclosure (3.7×7.4 m) with an adjacent haulout area [Fig. 1(a)], located in San Diego Bay. The study followed a protocol approved by the Institutional Animal Care and Use Committee at SPAWAR Systems Center, San Diego and followed all applicable U.S. Department of Defense guidelines. Both subjects were healthy during the course of the study.

B. Experimental apparatus

1. Underwater stations

The test apparatus consisted of two underwater listening stations, designated as the ''S1 station'' and the ''S2 station.'' The S1 station was the site for the presentation of a ''start'' signal to begin the hearing test and the underwater impulse produced by the PPD. The hearing tests were conducted at the S2 station. Two stations were used to spatially separate the location of the hearing test (the S2 station) from the site of exposure to the PPD impulsive sound (the S1 station). A similar approach was used with dolphins and white whales by Schlundt *et al.* (2000) and Finneran *et al.* (2000, 2002b). Each station consisted of a polyvinylchloride (PVC) frame with a plastic biteplate on which the subjects were trained to position. The S1 and S2 biteplates were located at depths of approximately 1.7 and 2 m, respectively. Each station contained an underwater video camera.

The S1 station [Fig. 1(b)] contained a single sound projector (ITC 1032) that was used to emit a 1-s tone as the start signal for the subject to begin hearing tests. These start tones, or ''S1 tones,'' were at a frequency of 6 kHz and SPL of 120 and 130 dB *re*: 1 μPa for NRT and LIB, respectively (LIB responded better to the higher level). The S1 start tones were produced using a personal computer (PC) with a National Instruments PCI-MIO-16E-1 multifunction board, then amplified (BGW PS2) and input to the ITC 1032.

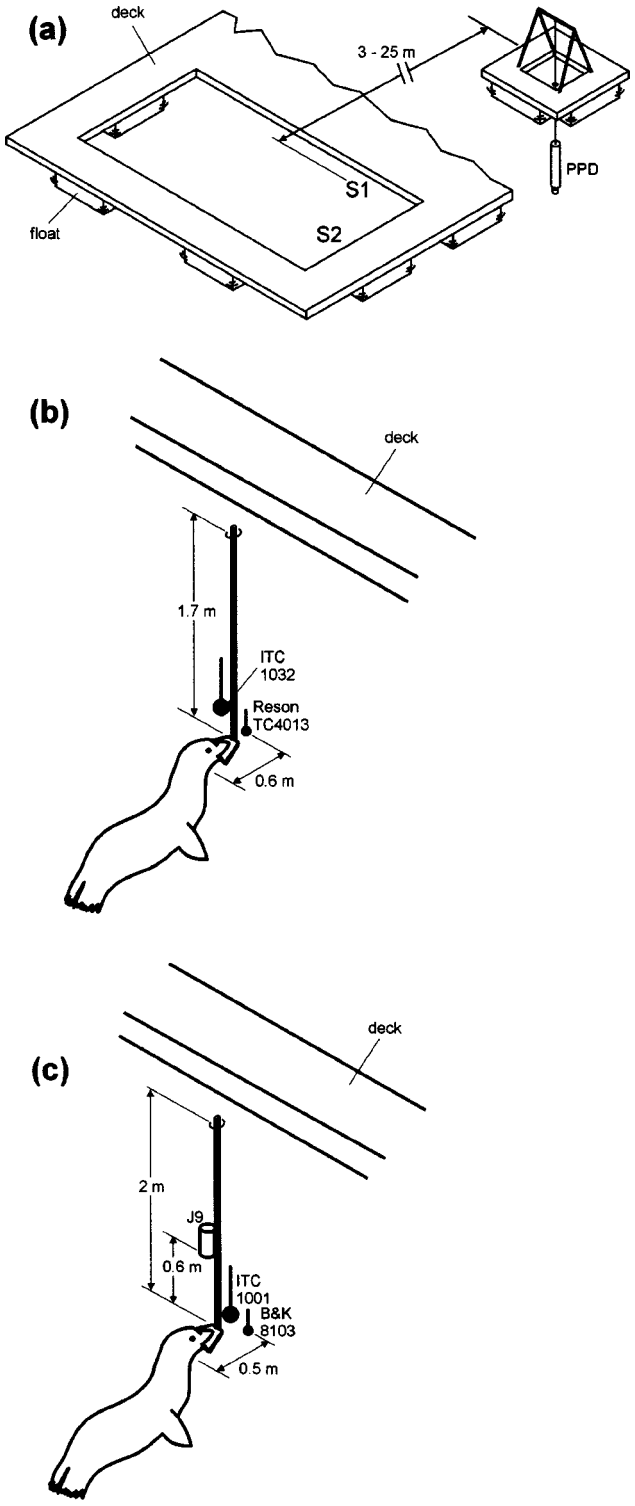


FIG. 1. (a) Experimental test site for the sound exposures and the hearing tests. (b) S1 station detail. (c) S2 station detail.

The S2 station [Fig. 1(c)] contained two sound projectors: one (ITC 1001) used to project hearing test tones, or ''S2 tones,'' and one (USRD J9) used to project masking noise. Hearing tests were performed at 1 and 10 kHz. The inclusion of additional test frequencies, while desirable, was not practical because of time constraints. The specific frequencies were selected to strike a balance between the range of frequencies present in the PPD waveform (see Sec.

III B 2), and the audible frequency range of *Zalophus* (see Schusterman *et al.*, 1972; Kastak and Schusterman, 1998). The order in which the S2 frequencies were tested was varied from day to day. The S2 tones were 250 ms in duration including 50-ms rise and fall times. S2 tones were generated using the PCI-MIO-16E-1, attenuated (HP 350D), filtered (Wavetek 452), and amplified (BGW PS4) before being input to the ITC 1001 sound projector.

Ambient noise levels at the test site (in San Diego Bay) were variable; thus, bandlimited white noise (masking noise) was introduced to keep thresholds consistent despite ambient noise fluctuations. Masking noise was generated on a second computer running custom software (see Finneran *et al.*, 2002a). This system continuously generated masking noise with a frequency spectrum that was compensated to eliminate the effects of projector frequency-dependent transmission characteristics and acoustic multipaths. Masking noise was generated using a PCI-MIO-16E-1, attenuated (HP 355D), filtered (Ithaco 4302), and amplified (BGW PS2) before being input to the USRD J9 sound projector. The projected masking noise was centered at the appropriate S2 frequency and had a frequency bandwidth equal to the test frequency; e.g., for hearing thresholds measured at 10 kHz, the noise frequency bandwidth extended from 5 to 15 kHz. The noise spectral density was 90 and 80 dB *re*: 1 μPa^2/Hz for S2 frequencies of 1 and 10 kHz, respectively. The noise spectral density was flat within ± 3 dB over these ranges. Noise was generated during the entire test period (i.e., throughout the time that the subject was in the test enclosure).

The acoustic pressure during each S2 tone presentation was measured using a B&K 8103 hydrophone (mounted to the S2 PVC frame), amplified (B&K 2635), filtered (SRS 560), digitized using the PCI-MIO-16E-1 multifunction board, and stored on the PC. The pressure during the S1 start tone presentation was measured using a Reson TC4013 hydrophone (mounted to the S1 PVC frame), amplified and filtered (B&K 2635, SRS 560), and digitized (PCI-MIO-16E-1). The PC was also used to record the time each S1 and S2 tone was produced.

2. Pulsed power device (PPD)

The PPD was developed and built by Pulse Power Technologies, Inc. In this device, a large electrical voltage is generated between two oppositely charged electrodes immersed in the water column. The terminals are then brought incrementally closer until an electrical arc bridges the small gap between them. The electrical arc momentarily vaporizes the water between the terminals, producing gas bubbles which quickly collapse. The expansion and collapse of the gas bubbles produces an impulsive pressure waveform in the water. The PPD featured several discrete settings for the amount of electrical energy to be discharged through the arc gap, ranging from 1.32 to 2.77 kJ.

Each exposure condition was assigned a nominal level, from N1 to N5 for NRT and L1 to L4 for LIB, in the order in which they were tested. Testing with NRT was complete before the first exposures were conducted with LIB. After preliminary examination of the results for NRT, we concluded that the exposure level N1, which featured a neoprene sheath to reduce the sparker output, was substantially below TTS-inducing levels, and therefore LIB was tested at only four exposure conditions. The particular PPD charge setting and distance to the subject were chosen in an attempt to match the exposures L1–L4 with N2–N5. Unfortunately, we were not able to achieve the same pressures and energy flux levels at L4 as we obtained at level N5. The PPD had a tendency to misfire (for an electrical arc to bridge the gap before intended) at the higher charge settings; for this reason level L4 was tested with 2.44-kJ charge instead of 2.77 kJ.

The acoustic pressure produced by the PPD was measured during each subject exposure using a hydrophone (Reson TC4013) mounted to the S1 PVC frame between the subject and the PPD; the hydrophone was located 0.6 m from the subject's ears. The hydrophone output was amplified and filtered from 2 Hz–100 kHz (B&K 2635 and SRS 560), then digitized at 200 kHz using the PCI-MIO-16E-1. The maximum pressure ($p_{\max}$), minimum pressure ($p_{\min}$), peak-to-peak SPL ($SPL_{\text{p-p}}$), rms SPL (SPL_{rms}), total energy flux (E_{T}), and effective duration (τ) were calculated from the digitized waveforms. A distance correction based on daily calibration measurements using a second hydrophone (B&K 8105 with a B&K 2635 amplifier) located at the location of the subjects' ears was used to correct the sound levels measured during exposure (0.6 m from the subjects' ears) to the actual received sound levels at the ear position.

The maximum and minimum pressures were defined as the maximum positive or negative peak, respectively, in the measured waveform. The effective duration of each impulse was defined as the difference between the times at which the cumulative integral of the instantaneous pressure squared $[p^2(t)]$ reached 5% and 95% of its final value (see Clark *et al.*, 1999). The energy flux spectral density $E(m)$ was calculated as

$$E(m) = |P(m)|^2, \quad m = 0,1,...,N-1, \tag{1}$$

where

$$P(m) = \Delta t \sum_{n=0}^{N-1} p(n) e^{-j2\pi mn/N}, \quad m = 0,1,...,N-1, \tag{2}$$

where $p(n)$ is the digitized pressure waveform, $P(m)$ is the discrete Fourier transform of $p(n)$, N is the number of samples in $p(n)$, Δt is the sampling interval, and $j = (-1)^{1/2}$. Fourier analysis of each signal was based on an 82-ms time window (to produce the desired frequency resolution). Note that the medium characteristic impedance ρc has been eliminated from the definition of $E(m)$ described by Fricke *et al.* (1985) and Johnston *et al.* (1988); this was done to emphasize that the energy flux spectral density was based on individual pressure measurements (rather than actual acoustic intensity measurements). The energy flux spectral density thus has units of μPa$^2\cdot$s/Hz, rather than J$\cdot$m$^{-2}\cdot$Hz^{-1} (0 dB *re*: 1 J$\cdot$m$^{-2}\cdot$Hz^{-1} would be equivalent to 182 dB *re*: 1 μPa$^2\cdot$s/Hz, for seawater with nominal values of density $\rho = 1026$ kg/m^3 and sound speed $c = 1500$ m/s). The energy flux density E_T was calculated from

TABLE I. Exposure conditions for NRT and LIB. Mean values are listed. The ± values indicate the standard deviation for linear units (e.g., kPa) and the maximum± excursion for values in dBs.

Subject	Level	Charge (kJ)	Distance (m)	p_{max} (kPa)	p_{min} (kPa)	$SPL_{p\text{-}p}$ (dB *re*: 1 μPa)	SPL_{rms} (dB *re*: 1 μPa)	E_T (dB *re*: 1 $\mu Pa^2 \cdot s$)	τ (ms)
NRT	N1	1.32[a]	15	0.065± 0.015	−0.041± 0.016	160± 2	143± 1	128± 1	28.3± 8.0
NRT	N2	1.32	25	0.49± 0.063	−0.74± 0.28	182± 1	155± 2	137± 1	16.2± 3.0
NRT	N3	1.32	6.5	1.2± 0.086	−2.2± 0.029	190± 1	169± 1	148± 1	10.5± 0.63
NRT	N4	2.44	6.5	7.0± 1.4	−6.3± 2.7	202± 2	177± 2	158± 2	13.9± 0.49
NRT	N5	2.77	3.4	14± 6.0	−4.9± 0.12	205± 3	183± 1	163± 1	14.2± 0.12
LIB	L1	1.32	25	0.54± 0.18	−0.96± 0.093	183± 1	153± 4	139± 1	18.7± 3.6
LIB	L2	2.19	15	1.7± 0.35	−1.4± 0.21	190± 1	165± 1	149± 2	18.3± 2.3
LIB	L3	2.19	6.5	5.3± 0.096	−5.8± 0.026	201± 1	175± 1	156± 1	13.8± 0.59
LIB	L4	2.44	3.4	6.8± 0.64	−3.3± 1.2	200± 1	178± 1	161± 4	11.5± 0.60

[a]At level N1 a neoprene rubber sheath was placed over the electrode assembly to reduce the sound output.

$$E_T = \Delta t \sum_{n=0}^{N-1} p^2(n), \tag{3}$$

where ρc was again removed from the equation. The energy flux density has units of $\mu Pa^2 \cdot s$ (0 dB *re*: 1 J/m^2 would be equivalent to 182 dB *re*: 1 $\mu Pa^2 \cdot s$, for seawater with $\rho = 1026$ kg/m^3 and $c = 1500$ m/s).

Table I shows the mean values for p_{max}, p_{min}, $SPL_{p\text{-}p}$, SPL_{rms}, E_T, and τ at each exposure condition for NRT and LIB. The values following the ± symbols in Table I indicate the standard deviations for linear quantities and the maximum± excursions for decibel quantities. Figures 2 and 3 show representative pressure waveforms and energy flux spectral densities measured for each of the exposure conditions for NRT and LIB, respectively. Note that the ordinate and abscissa scales are not uniform throughout the panels. Pressure waveforms in Figs. 2 and 3 show the effects of multipath propagation caused by reflections from the surface and bottom. Many of the recorded waveforms show two positive pressure peaks separated by several milliseconds, presumably caused by bottom reflection. This second pressure peak resulted in higher total energy fluxes than one would have encountered in open water without the bottom-reflected component.

C. Procedure

1. Overview

Figure 4 illustrates the test procedure and hearing threshold calculation technique. The test procedure was similar to that used by Finneran *et al.* (2000, 2002b). Each subject participated in one test session each day. Each test session was either an exposure session, where the subject was exposed to a single underwater impulse produced by the PPD, or a control session, where the PPD firing sequence was simulated but no impulse was produced (a ''mock'' exposure). During each test session, three hearing tests were performed [Fig. 4(a)]. Two hearing tests were conducted before the exposure (or mock exposure); these tests produced preexposure hearing thresholds at the two S2 frequencies (1 and 10 kHz). One hearing test was performed after the exposure (postexposure). Preexposure thresholds were measured at both 1 and 10 kHz to verify that the subject's hearing was within normal baseline values at each frequency. The order in which the preexposure frequencies were tested was varied from day to day. The postexposure hearing test was performed at the last S2 frequency tested preexposure. Table I lists the individual exposure conditions for NRT and LIB. Each exposure condition was repeated over 2 days to conduct hearing tests at each S2 frequency immediately before and after exposure. Eight control sessions were conducted for each subject: four control sessions with 1 kHz tested postexposure and four control sessions with 10 kHz tested postexposure. The control sessions were randomly interspersed within the exposure sequence.

The preexposure and postexposure hearing test procedures were very similar. The preexposure hearing test procedure is presented in detail, followed by a discussion of the postexposure hearing test procedure.

2. Preexposure hearing tests

The hearing test procedure was based on the method of free response, or MFR (Egan *et al.*, 1961). The hearing tests were divided into a number of relatively long observation periods, called ''dives.'' Each dive began with the trainer directing the subject (with a hand signal and spoken command) to the S1 station. The subject was trained to remain on the S1 station until presented with the S1 start tone. Upon hearing the S1 start tone, the subject proceeded to the S2 station. Once the subject was positioned at the S2 station, a block of hearing test trials was presented. The trial block was ended when the trainer sounded an underwater buzzer to signal the subject to return to the surface and receive fish reward. The process was then repeated as necessary.

Each trial block contained a variable number of trials. Each trial was 2.0 s in duration. The time interval between trials (the interstimulus interval, or ISI) was randomized between 5 and 12 s. Neither the subject nor the trainer knew when the next trial would occur. Eighty percent of the trials contained a 250-ms S2 tone beginning 50 ms before the trial start. The remaining 20% of the trials were no-tone or catch trials. Since the subject did not know when the next trial would occur, the catch trial periods functioned as ''equipment catch trials'' and were primarily used to confirm that the sound system was not producing artifacts coinciding with the stimulus.

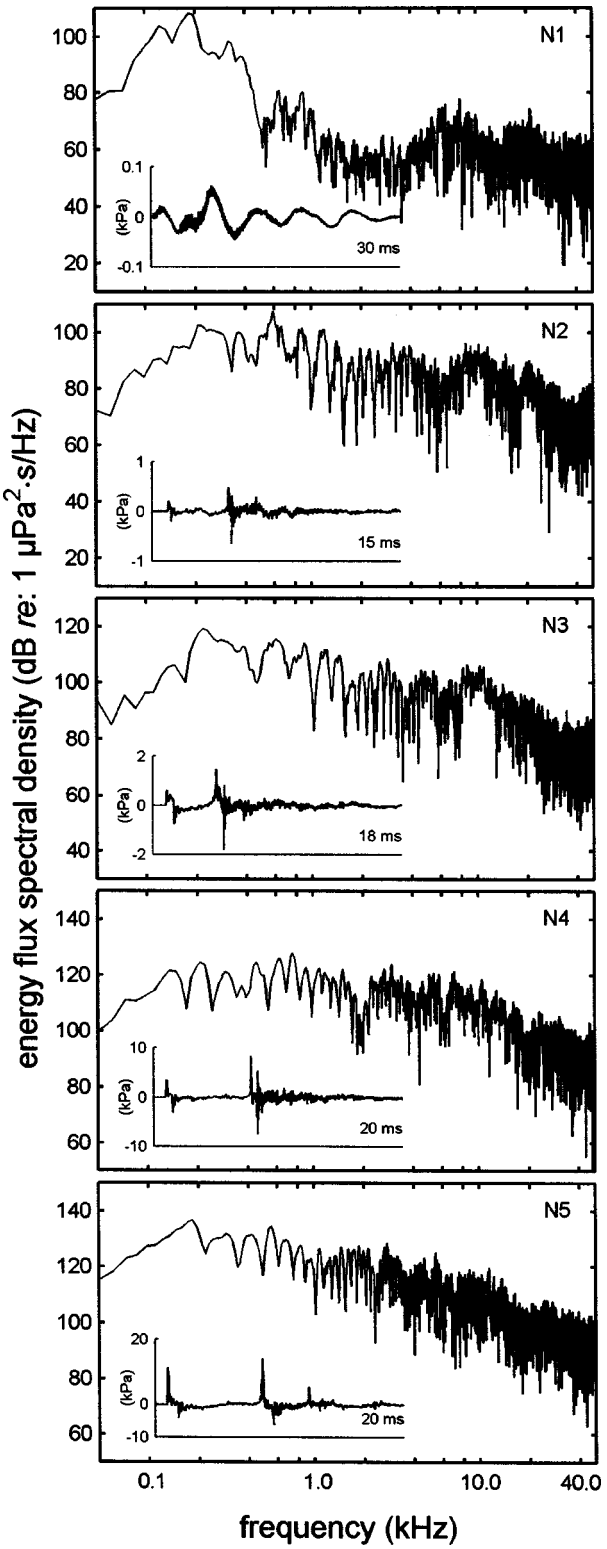


FIG. 2. Representative pressure waveforms and energy flux spectral densities measured for exposure conditions N1–N5. Note that the ordinate and abscissa scales are not uniform throughout the panels.

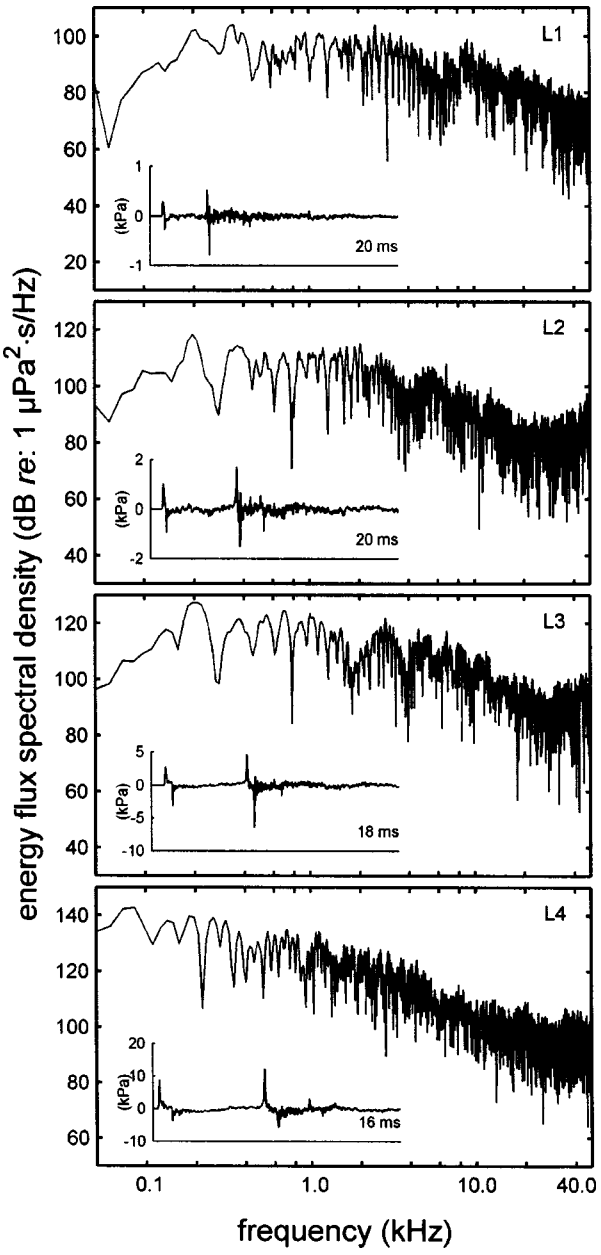


FIG. 3. Representative pressure waveforms and energy flux spectral densities measured for exposure conditions L1–L4. Note that the ordinate and abscissa scales are not uniform throughout the panels.

Subjects were trained to produce a vocal response if they heard a tone and to remain quiet otherwise. Each tone trial had two possible outcomes: Vocal responses occurring within a tone trial were recorded as ''hits.'' No vocal response during a tone trial was a ''miss.'' The amplitudes of the S2 tones were adjusted using a modified up/down staircase procedure (e.g., Cornsweet, 1962): the amplitude was decreased 2 dB following each hit and increased 2 dB following each miss.

The trainer and computer operator monitored the sound in the water (using the Reson TC4013 hydrophone mounted on the S2 frame) for any vocal responses by the subject. The computer operator and trainer were aware of the session type (control or exposure), the type of trial (catch or tone), the S2 tone amplitude during each tone trial, and were notified as each trial began. The point at which reinforcement was delivered (i.e., the number of trials per trial block) was determined by the computer operator. The number of trials per trial block was randomly varied within the following guidelines: Only correct responses were reinforced. An attempt was made to reinforce responses to low-level tones (i.e., at a lower level than any previously responded to). If the subject missed several tones in a row, the first subsequent hit was

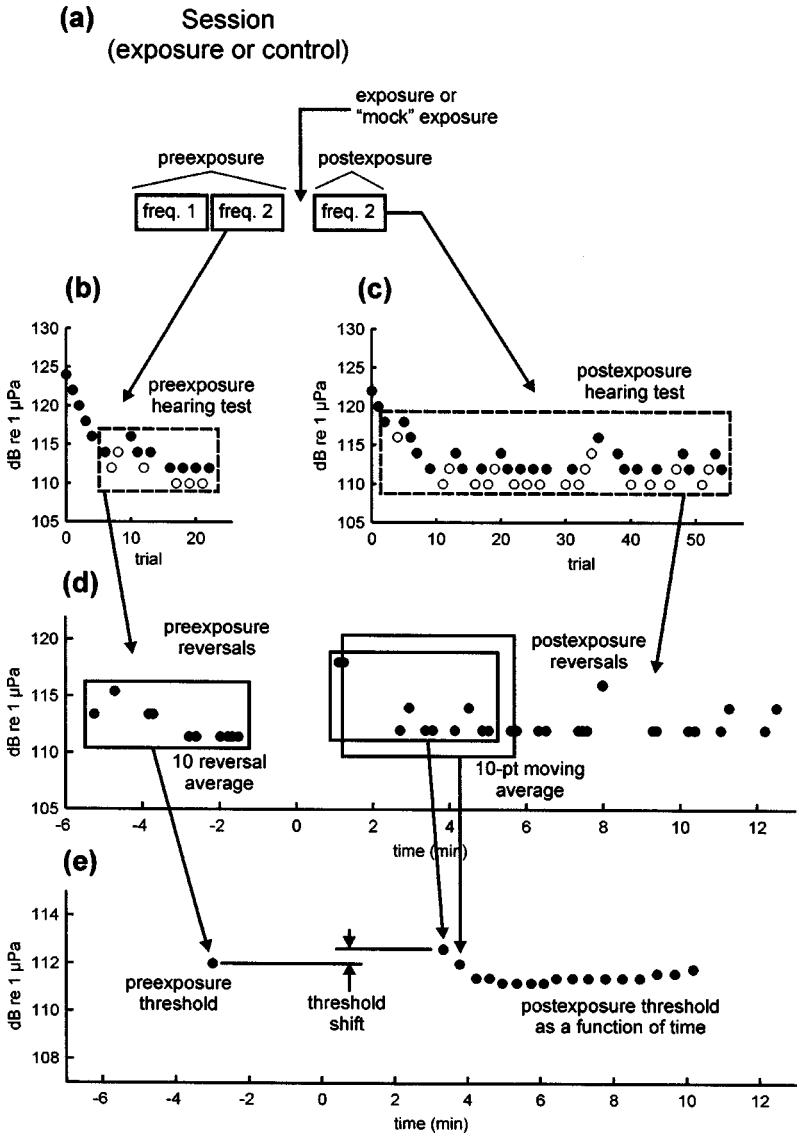


FIG. 4. Test sequence and threshold calculation method. (a) A single test session (exposure or control) was conducted each day. During each session, three hearing tests were conducted: two preexposure and one postexposure. (b) Preexposure and (c) postexposure hearing test results consisted of the subjects' performance (hit or miss) as a function of the S2 tone SPL. The filled and open symbols represent hits and misses, respectively. (d) Preexposure thresholds were calculated from the average pressure over the first ten reversals. Postexposure thresholds were estimated using a ten-pt moving average. (e) The threshold shift was determined by subtracting the preexposure thresholds from the postexposure threshold.

generally not reinforced. The dive times were normally kept under 2 min. The amount of reinforcement was scaled to the performance of the subject during the dive (e.g., more reinforcement was given for longer dives and/or responding to low-level tones).

After reinforcement, the next dive was begun and the procedure repeated until the hearing test was complete. Preexposure hearing tests were conducted until at least ten ''reversals,'' defined as a transition from a hit to a miss or from a miss to a hit, were obtained. The first S2 tone at each frequency was presented approximately 10 dB above the expected threshold. Each preexposure threshold could usually be estimated after 20 to 30 trials.

Figures 4(b)–(e) illustrate the threshold calculation method. The preexposure hearing test resulted in a record of the subject's performance (hit or miss) to each tone trial [Fig. 4(b)]. The performance data were then converted to a series of reversals, shown in Fig. 4(d). The time and amplitude for each reversal were defined as the mean time and mean pressure, respectively, of the hit/miss pair. Time values in Fig. 4(d) are referenced to the time of the exposure (or mock exposure); thus, the preexposure reversals occurred at negative times. Finally, preexposure hearing thresholds were calculated as the mean pressure of the first ten hit–miss/miss–hit reversal points [Fig. 4(e)]. The time at which the threshold occurred was defined as the mean time of the ten reversals.

3. Postexposure hearing tests

The postexposure hearing test procedure was identical to the preexposure procedure with two exceptions: (1) During exposure sessions, a single impulse was produced by the PPD approximately 0.5 s before the start of the first S1 start tone of the postexposure hearing test. During control sessions, a switch was used to disable the PPD triggering signal, preventing the PPD from firing. (2) The postexposure hearing tests were conducted for at least 10 min to enable any TS and recovery to be tracked.

Figures 4(c)–(e) illustrate the postexposure threshold calculation. As with the preexposure tests, the postexposure hearing test resulted in a record of the subject's performance (hit or miss) to each tone [Fig. 4(c)], which was converted to a series of reversals, shown in Fig. 4(d). The time and am-

TABLE II. Mean values for preexposure masked hearing thresholds and false alarm rates for NRT and LIB. The ± values are standard deviations. Data from exposure and control sessions are pooled.

	NRT			LIB		
Frequency (kHz)	Threshold (dB *re*: 1 μPa)	R_{FA} (%)	*n*	Threshold (dB *re*: 1 μPa)	R_{FA} (%)	*n*
1	118±2	5.3±2.4	18	122±2	3.3±3.1	16
10	115±4	5.5±3.2	18	114±2	3.4±2.8	16

plitude for each reversal were defined as the mean time and mean pressure, respectively, of the hit/miss pair. Since the time values in Fig. 4(d) are referenced to the time of the exposure (or mock exposure), the postexposure reversals occurred at positive times. Postexposure hearing thresholds were calculated by applying a ten-point moving average to the postexposure reversals [Figs. 4(d)–(e)]. Each output of the moving average consisted of two values: the mean pressure over the ten reversals and the mean time over which the ten reversals occurred. The postexposure threshold as a function of time thus represented the mean time and mean pressure of each group of ten postexposure reversals. Thresholds at specific postexposure times (e.g., 5 min, 10 min) were obtained by interpolating within the collection of thresholds from the moving average.

The TS was calculated by subtracting the preexposure threshold from the postexposure threshold. An MTTS was defined as a 6-dB or larger TS. This 6-dB criterion was based on other marine mammal TTS studies using similar test paradigms (e.g., Schlundt *et al.*, 2000; Finneran *et al.*, 2000) and was considered to be the minimum shift that was larger than any day-to-day or session-to-session variations in the subjects' masked hearing thresholds.

By definition a TTS will recover after some amount of time (i.e., the postexposure threshold will return to the preexposure level). Meaningful comparisons of TSs resulting from different exposure conditions thus require the TSs to be measured at the same time postexposure. The first postexposure threshold was often obtained (i.e., the mean time of the first ten postexposure reversals was) about 3–5 min after the exposure. Thresholds were obtained within 5-min postexposure in nearly all (33/34; the exception was 5.6-min postexposure) cases; thus, the threshold shifts at approximately 5 min (TS_5) and 10 min (TS_{10}) postexposure were convenient metrics to compare the effects of different exposures.

4. False alarm rate

Any vocal response during a catch trial was recorded as a false alarm. Any vocal response by a subject not occurring within a trial was also recorded as a false alarm. The ISI (defined from the start of one tone to the start of the next tone) was randomly varied between 5–12 s; thus, the majority of time spent on the S2 station was outside any trials and functioned as an additional ''catch trial'' period.

The false alarm rate R_{FA} was defined as

$$R_{FA}=\frac{N_{FA}}{T-N_{S2}T_1}T_1, \tag{4}$$

where N_{FA} is the number of false alarms, T is the total amount of time the subject spent on the S2 station, N_{S2} is the number of S2 tone trials presented, and T_1 is the trial duration. For the MFR, R_{FA} values calculated using Eq. (4) are analogous to false alarm rates obtained from a single interval experiment (Miller, 1969); however, this study employed a modified version of the MFR: the ISI was not open-ended but randomized between 5 and 12 s. The R_{FA} values calculated here are therefore not identical to those obtained with the MFR or to false alarm rates obtained from a single interval experiment; however, they do reliably assess a subject's response bias from session to session. The denominator of Eq. (4) is the total amount of time during which the subject was on the S2 station without a tone trial present; therefore, R_{FA} is a way of normalizing the number of false alarms with respect to the amount of time that the subject had an opportunity to commit a false alarm (see Finneran *et al.*, 2002a; 2002b).

III. RESULTS AND DISCUSSION

A. PPD waveforms

The shallow test site depth affected the pressure waveforms produced at the location of the subject. The most striking feature was the presence of a second positive peak in the pressure waveform, presumably caused by the bottom-reflected wave from the PPD. The presence of the bottom reflected wave resulted in higher energy flux densities than would exist in PPD waveforms in open water. The differences between the pressure waveforms measured in the current study and those expected from the PPD in open water illustrate the importance of basing impact criteria for particular effects (such as TTS) on the actual received levels that produced the effect, rather than, for the case of the PPD, electrical charge and distance combinations that produced the effects.

B. Preexposure thresholds and false alarm rates

Table II shows the preexposure masked hearing thresholds and false alarm rates measured for NRT and LIB. In this table, data from both control and exposure sessions are pooled. Masked thresholds and false alarm rates for NRT at 1 and 10 kHz were 118 and 115 dB *re*: 1 μPa and 5.3% and 5.5%, respectively. Masked thresholds and false alarm rates for LIB at 1 and 10 kHz were 122 and 114 dB *re*: 1 μPa and 3.3% and 3.4%, respectively. False alarm rates for both subjects had large variances (standard deviations were of the same order of magnitude as the mean values). Thresholds were generally within ±4 dB from day-to-day. The variability in the preexposure hearing thresholds was similar to that

TABLE III. Mean values for postexposure false alarm rates for NRT and LIB. The ± values are standard deviations.

Frequency (kHz)	NRT				LIB			
	Control		Exposure		Control		Exposure	
	R_{FA} (%)	n	R_{FA} (%)	n	R_{FA} (%)	n	R_{FA} (%)	n
1	5.3±1.3	4	6.2±4.7	5	3.2±3.1	4	4.6±3.2	4
10	6.7±4.5	4	4.1±3.6	5	4.5±4.4	4	2.4±0.4	4

observed in behavioral hearing tests with human and animal subjects (e.g., Johnson, 1967; NIOSH, 1998).

The critical ratio (CR) is defined as the ratio of the masked threshold to the noise power spectral density (Fletcher, 1940). CRs are a measure of the auditory system's ability to detect a signal in noise, and may be used to compare masked hearing thresholds measured with different noise levels. CRs for the present study were 28 and 32 dB *re*: 1 Hz and 35 and 34 dB *re*: 1 Hz at 1 and 10 kHz, respectively. Masked thresholds and CRs at 1 kHz were similar to those measured in *Zalophus* by Southall *et al.* (2000). Southall *et al.* reported thresholds of 122 and 119 dB *re*: 1 μPa and CRs of 28 and 22 dB *re*: 1 Hz at 800 and 1200 Hz, respectively. Southall *et al.* used octave-band masking noise spectral density levels of 94 and 97 dB *re*: 1 μPa^2/Hz at 800 and 1200 Hz, respectively (at 1 kHz, an octave band would extend from approximately 700–1400 Hz, which is smaller than the bandwidth employed in the present study). There are no existing data for CRs above 2.5 kHz in *Zalophus*. CRs have been measured near 10 kHz in the harbor seal (20 dB *re*: 1 Hz at 8 kHz; Terhune, 1991; Turnbull and Terhune, 1990) , northern fur seal, *Callorhinus ursinus* (21 dB *re*: 1 Hz at 8 kHz; Moore and Schusterman, 1987), ringed seal, *Phoca hispida*, (32 dB *re*: 1 Hz at 8 kHz; Terhune and Ronald, 1975), and harp seal, *Phoca groenlandica*, (35 dB *re*: 1 Hz at 8.6 kHz; Terhune and Ronald, 1971). The current data are at the high end of this range. Unfortunately, there are insufficient data to identify whether the measured CRs are within the normal range of variability for *Zalophus*, or if they indicate a loss of hearing sensitivity at 10 kHz. Any preexisting hearing loss at 10 kHz would have resulted in less observed TS than one would measure from an unimpaired listener.

Table III compares the postexposure false alarm rates measured during control and exposure sessions. As with the preexposure data, variances in the false alarm rates were relatively high. There were no significant differences between the control and exposure session mean false alarm rates for either NRT [$t(5)=0.398$, $p=0.707$ at 1 kHz and $t(6)=-0.952$, $p=0.380$ at 10 kHz] or LIB [$t(5)=0.421$, $p=0.692$ at 1 kHz and $t(3)=-0.952$, $p=0.411$ at 10 kHz].

C. Direct effects of the exposures

Neither subject showed any signs of physical injury as a result of exposure to the impulsive sounds produced by the PPD. There were no 6-dB or larger threshold shifts observed (i.e., no MTTS according to the 6-dB criterion) in any of the exposure or control sessions. At the conclusion of the study both subjects' hearing thresholds were within normal baseline ranges; there were no permanent auditory, appetite, or health effects of the exposures.

Figure 5 shows the threshold shifts measured at 5 min (TS_5) and 10 min (TS_{10}) postexposure plotted versus the exposure condition for NRT [Fig. 5(a)] and LIB [Fig. 5(b)]. The upper and lower panels in each figure show the results obtained at 1 and 10 kHz, respectively. Filled symbols show the amount of TS_5; open symbols show the measured TS_{10}. The data points for TS_5 and TS_{10} are offset for clarity. The control session results are grouped together. Figure 5 provides a way of comparing the amounts of TS obtained after exposures to the amount of TS observed during control sessions. There are no obvious relationships between the expo-

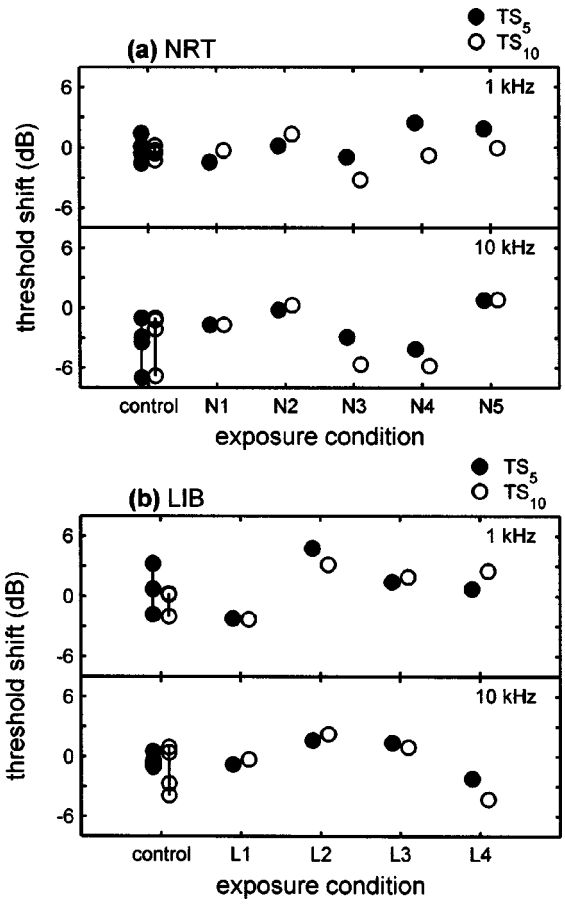


FIG. 5. Threshold shifts measured 5 min (TS_5, filled symbols) and 10 min (TS_{10}, open symbols) after exposure for (a) NRT and (b) LIB. The upper panel in each figure shows the results at 1 kHz, the lower panel shows the results at 10 kHz. Results from control sessions are grouped. The NRT, 10-kHz, control, TS_5 data point at −7 dB was obtained at 5.6-min postexposure.

sure condition and the resulting TS in the data of Fig. 5; for these two subjects the particular exposure conditions used were below those necessary to cause a significant MTTS. Most (31/36 for NRT, 30/32 for LIB) of the measured TSs were within ±4 dB. All of the control sessions and most of the exposure sessions resulted in negative amounts of TS for NRT at 10 kHz. Since this occurred in all of the control sessions, it seems likely that this is an artifact peculiar to this subject at 10 kHz with the test paradigm employed. The TS results for NRT at 1 kHz and for LIB were distributed more evenly about zero. The values of TS_5 and TS_{10} were normally within ±2 dB of each other for a particular test. In summary, the TSs observed and differences between TS_5 and TS_{10} were within the normal variability associated with behavioral hearing tests and should not be interpreted as significant.

D. Indirect (behavioral) effects of the exposures

Previous marine mammal TTS studies have reported ''behavioral reactions'' in the test subjects after exposure to high levels of underwater sound (Kastak *et al.*, 1999; Schlundt *et al.*, 2000). These reactions mainly consisted of alterations in the subjects' trained behaviors apparently designed to allow the subject to avoid the sound exposure itself (for long duration exposures) or to avoid the spatial location of the exposure during subsequent tests. More severe reactions such as aggression toward the test apparatus have also been observed (e.g., Schlundt *et al.*, 2000). In the current study, alterations in the trained behaviors of NRT and LIB began to occur as the exposure levels increased.

The reaction of NRT was mild: At level N3 NRT did not leave the S1 station after exposure to the PPD impulse and the S1 start signal, and required a second S1 start signal to be presented before he proceeded to the S2 station. At levels N4 and N5, the subject returned to the trainer after exposure to the impulsive sound. Following this, the trainer directed NRT to the S1 station, another S1 start signal was presented, and the subject proceeded to the S2 station for the hearing test. It is unknown why NRT returned to the trainer after exposure to the impulse. It is possible that he initially confused the impulsive sound with a commonly used ''recall'' sound produced by the trainer slapping the water surface.

The reactions of LIB were more substantial. At levels L2, L3, and L4 LIB was reluctant to return to the S1 station after exposure to the PPD impulse. The severity of this behavior (i.e., the amount of time required for him to resume stationing on the S1 biteplate) increased as the test sequence progressed. As the testing progressed over a number of days, LIB became reluctant to station on the S1 biteplate *before* the exposure. Other reactions included looking around with his head out of the water (level L3) and hauling out (level L4).

It is important to distinguish these deviations in the subjects' trained behaviors from reactions that may occur in wild and/or naive animals. The exposure paradigm, with exposure levels increasing over a number of days, likely had an impact on the subjects' behaviors (i.e., the subjects knew that the exposure levels were increasing from day to day). [This particular exposure sequence was a conservative approach used because the PPD was a novel stimulus. If significant amounts of TTS had been observed after an exposure, testing would not have continued to higher levels.] The fact that one subject (LIB) was reluctant to station on the S1 biteplate before the exposure indicates that the subjects' prior experiences with the test apparatus also affected their reactions. Although the subjects both showed clear behavioral reactions to the impulsive sounds, no changes in general health or appetite were detected.

E. Effects of masking noise

There have been few studies of the effects of masking noise on the amount of observed TS in mammals. Parker *et al.* (1976) and Humes (1980) showed that the presence of masking noise resulted in elevated hearing thresholds (simulating a preexposure loss in hearing sensitivity) and decreased the amount of TTS measured in humans. Ades *et al.* (1974) also measured smaller amounts of PTS in chinchillas when thresholds were obtained in the presence of masking noise compared to the amount of PTS measured when thresholds were determined in quiet. Unfortunately, there are no conclusive data for the relationship between masking noise and TTS in marine mammals (see Schlundt *et al.*, 2000); it is possible that larger TSs may have been observed in the present study if the masking noise was not used. Masking noise was employed because the test site had a variable ambient noise level. The masking noise level was the lowest at which uniform frequency content and stable hearing thresholds could be maintained. Preexposure hearing thresholds (Table II) were approximately 30–35 dB above published thresholds for *Zalophus* (Schusterman *et al.*, 1972; Kastak and Schusterman, 1998); however, subjects' thresholds measured in San Diego Bay without masking noise (ambient noise levels in San Diego Bay at the test site were approximately 78 and 69 dB re: 1 μPa^2/Hz at 1 and 10 kHz, respectively) were only approximately 7–14 dB below the masked thresholds.

F. Comparison to existing data

Kastak *et al.* (1999) measured 2.9–6.7 dB of TTS in a California sea lion exposed to 20 min of octave band noise (1 and 2 kHz noise center frequencies). Figure 6 compares the TTS-inducing stimuli used by Kastak *et al.* to the exposure conditions used in the current study. Figures 6(a) and (b) show the peak pressure and rms pressure of each exposure, respectively, versus the exposure duration; Fig. 6(c) plots the exposure energy flux versus the duration. The open circles represent the data from the current study (no MTTS) and the solid rectangles show the range of exposures used by Kastak *et al.* (1999). Peak pressures for the Kastak *et al.* data were approximated as the rms pressure +3 dB.

Figures 6(a) and (b) illustrate the importance of considering both amplitude and duration when estimating the impact of sounds on marine mammals: Although some of the PPD impulses had relatively high peak and rms pressures (higher than those causing TTS in the Kastak *et al.* study), the duration of these impulses was relatively short; thus, the total energy was much lower than the energy in the noise exposures used by Kastak *et al.* [Fig. 6(c)]. Thus, it is not

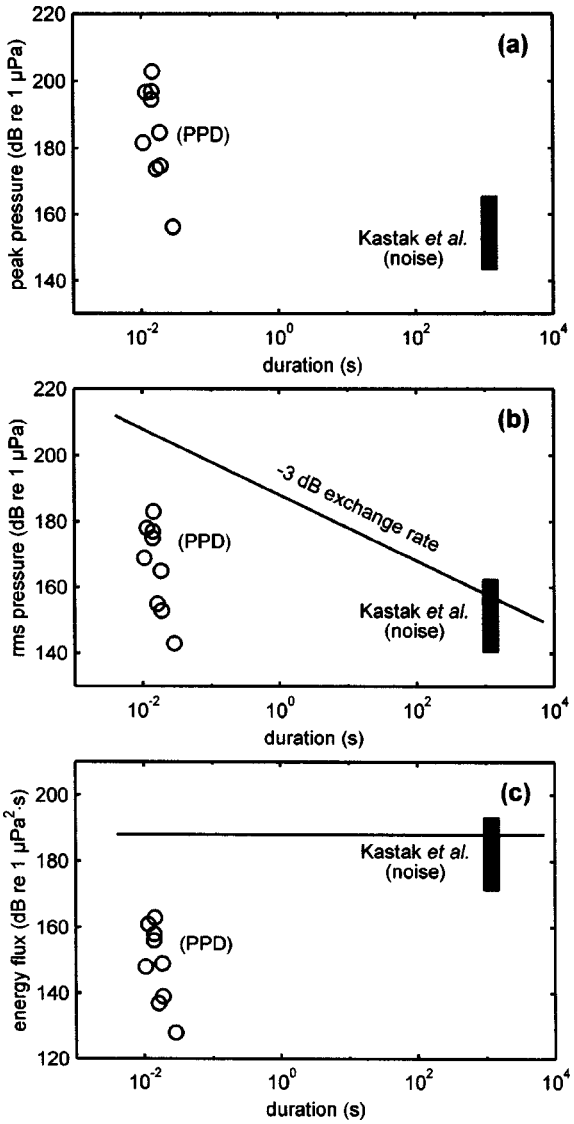


FIG. 6. Comparison of the PPD exposure conditions to those of Kastak *et al.* (1999). Exposure conditions are displayed as the (a) peak pressure; (b) rms pressure; and (c) total energy flux versus the exposure duration. The line labeled ''−3-dB exchange rate'' has a slope of −3 dB per doubling of time and is equivalent to an equal energy criterion for continuous-type sounds.

surprising that the single impulses from the PPD (at the charge levels and distances employed) did not cause a TTS in the subjects studied.

The solid line in Fig. 6(b) is a line with a slope of −3 dB per doubling of time fit to the mean value of the Kastak *et al.* data set. This is sometimes called a ''−3-dB exchange rate'' or an ''equal-energy criterion,'' since any two continuous-type sounds whose rms pressures and durations fall on the line will have the same total energy flux. For single, continuous exposures, exposures of equal energy lead to approximately equal effects on mammalian auditory systems (Ward, 1997). The limited amount of odontocete onset-TTS data from single impulsive, tonal, and noise exposures are also fit reasonably well by the −3-dB exchange rate (Finneran *et al.*, 2002b). It therefore seems reasonable that an equal energy rule would apply to pinnipeds as well (and thus the Kastak *et al.* data could be extrapolated to other durations on an equal energy basis); however, this has yet to be experimentally verified (see Southall *et al.*, 2001).

IV. CONCLUSIONS

(1) Auditory and behavioral effects of the PPD were assessed by measuring hearing thresholds in two trained California sea lions before and after exposure to single underwater impulses produced by the PPD.
(2) The pressure waveforms produced by the PPD were strongly influenced by the shallow depth at the test site. Most received pressure waveforms featured a second positive pressure peak produced from the bottom reflection. This condition resulted in larger total energy fluxes than one would have in open water without a bottom-reflected wave.
(3) No MTTS was observed in either subject at the highest received levels: peak pressures of approximately 6.8 and 14 kPa, rms pressures of approximately 178 and 183 dB re: 1 μPa, and total energy fluxes of 161 and 163 dB re: 1 μPa^{2}s for the two subjects. The total energy fluxes in the exposures were at least 25 dB lower than the mean energy flux level employed by Kastak *et al.* (1999), who observed small amounts of TTS in a California sea lion exposed to 20 min of octave band noise.
(4) Behavioral reactions to the tests were observed in both subjects. These reactions primarily consisted of attempts to avoid the site where exposure to the PPD impulse had previously occurred.

ACKNOWLEDGMENTS

The PPD was developed and maintained by Pulse Power Technologies, Inc. (PPTI) in Spring Valley, CA. Subjects were trained by Richard Jauck, Linda Green, Taylor Aguayo, Fred Shafer, Jennifer Jeffress, and Tricia Kamolnick. Wesley Elsberry and Carolyn Schlundt assisted with data collection. We thank Dr. Robert Gisiner at the U.S. Office of Naval Research (ONR) and Christina Fahy at NOAA Fisheries for encouragement and support during the course of this study. We also thank two anonymous reviewers for helpful comments. Funding was provided by the Pacific States Marine Fisheries Commission (PSMFC), NOAA Fisheries, and ONR Marine Mammal S&T Program.

Ades, H. W., Trahiotis, C., Kokko-Cunningham, A., and Averbuch, A. (**1974**). ''Comparison of hearing thresholds and morphological changes in the chinchilla after exposure to 4 kHz tones,'' Acta Oto-Laryngol. **78**, 192–206.

Clark, J. A., Young, J. A., and Gaspin, J. B. (**1999**). ''Validation of a system for synthesizing distant signatures of underwater explosions for sea mammal hearing studies,'' Proc. Joint EAA/ASA meeting, Berlin, Germany, 14–19, March 1999.

Cornsweet, T. N. (**1962**). ''The staircase method in psychophysics,'' Am. J. Psychol. **75**, 485–491.

Egan, J. P., Greenberg, G. Z., and Schulman, A. I. (**1961**). ''Operating characteristics, signal detectability, and the method of free response,'' J. Acoust. Soc. Am. **33**, 993–1007.

Finneran, J. J., Schlundt, C. E., Carder, D. A., and Ridgway, S. H. (**2002a**). ''Auditory filter shapes for the bottlenose dolphin (*Tursiops truncatus*) and the white whale (*Delphinapterus leucas*) derived with notched-noise,'' J. Acoust. Soc. Am. **112**, 322–328.

Finneran, J. J., Schlundt, C. E., Dear, R., Carder, D. A., and Ridgway, S. H. (**2002b**). ''Temporary shift in masked hearing thresholds (MTTS) in odontocetes after exposure to single underwater impulses from a seismic watergun,'' J. Acoust. Soc. Am. **111** 2929–2940.

Finneran, J. J., Schlundt, C. E., Carder, D. A., Clark, J. A., Young, J. A., Gaspin, J. B., and Ridgway, S. H. (**2000**). ''Auditory and behavioral responses of bottlenose dolphins (*Tursiops truncatus*) and white whales (*Delphinapterus leucas*) to impulsive sounds resembling distant signatures of underwater explosions,'' J. Acoust. Soc. Am. **108**, 417–431.

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–61.

Fricke, J. R., Davis, J. M., and Reed, D. H. (**1985**). ''A standard quantitative calibration procedure for marine seismic sources,'' Geophysics **50**(10), 1525–1532.

Humes, L. E. (**1980**). ''Temporary threshold shift for masked pure tones,'' Audiology **19**, 335–345.

Johnson, C. S. (**1967**). ''Sound detection thresholds in marine mammals,'' in *Marine Bio-Acoustics*, edited by W. N. Tavolga (Pergamon, New York), Vol. 2, pp. 247–260.

Johnston, R. C., Reed, D. H., and Desler, J. F. (**1988**). ''Special report of the SEG technical standards committee, SEG standards for specifying marine seismic energy sources,'' Geophysics **53**(4), 566–575.

Kastak, D., and Schusterman, R. J. (**1996**). ''Temporary threshold shift in a harbor seal (*Phoca vitulina*),'' J. Acoust. Soc. Am. **100**, 1905–1908.

Kastak, D., and Schusterman, R. J. (**1998**). ''Low-frequency amphibious hearing in pinnipeds: Methods, measurements, noise, and ecology,'' J. Acoust. Soc. Am. **103**, 2216–2228.

Kastak, D., Schusterman, R. J., Southall, B. L., and Reichmuth, C. J. (**1999**). ''Underwater temporary threshold shift induced by octave-band noise in three species of pinniped,'' J. Acoust. Soc. Am. **106**(2), 1142–1148.

Miller, H., (**1969**). ''The FROC curve: A representation of the observer's performance for the method of free response,'' J. Acoust. Soc. Am. **46**, 1473–1476.

Moore, P. W. B., and Schusterman, R. J. (**1987**). ''Audiometric assessment of northern fur seals, *Callorhinus ursinus*,'' Marine Mammal Sci. **3**, 31–53.

NIOSH (**1998**). *Criteria for a Recommended Standard: Occupational Noise Exposure, Revised Criteria 1998*, DHHS (NIOSH) Publication No. 98–126 (NIOSH, Cincinnati, OH).

National Marine Fisheries Service (NMFS). (**1997**). *Investigation of Scientific Information on the Impacts of California Sea Lions and Pacific Harbor Seals on Salmonids and on the Coastal Ecosystems of Washington, Oregon, and California*, NOAA Tech. Memo. NMFS-NWFSC-28 (U.S. Dept. Commerce, Washington, DC).

Parker, D. E., Tubbs, R. L., Johnston, P. A., and Johnston, L. S. (**1976**). ''Influence of auditory fatigue on masked pure-tone thresholds,'' J. Acoust. Soc. Am. **60**, 881–885.

Reeves, R. R., Hofman, R. J., Silber, G. K., and Wilkinson, D. (**1996**). *Acoustic deterrence of harmful marine mammal-fishery interactions: proceedings of a workshop held in Seattle, Washington, 20–22 March, 1996*, NOAA Tech. Memo. NMFS-OPR-10 (U.S. Dept. Commerce, Washington, DC).

Richardson, W. J., Greene, C. R., Jr., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Schlundt, C. E., Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2000**). ''Temporary shift in masked hearing thresholds (MTTS) of bottlenose dolphins, *Tursiops truncatus*, and white whales, *Delphinapterus leucas*, after exposure to intense tones,'' J. Acoust. Soc. Am. **107**, 3496–3508.

Schusterman, R. J., Balliet, R. F., and Nixon, J. (**1972**). ''Underwater audiogram of the California sea lion by the conditioned vocalization technique,'' J. Exp. Anim. Behav. **17**, 339–350.

Shaughnessy, P. D., Semmelink, A., Cooper, J., and Frost, P. G. H. (**1981**). ''Attempts to develop acoustic methods of keeping Cape fur seals *Arctocephalus pusillus* from fishing nets,'' Biol. Conserv. **21**, 141–158.

Southall, B. L., Schusterman, R. J., Kastak, D., Reichmuth Kastak, C., and Holt, H. M. (**2001**). ''Pinniped hearing and anthropogenic noise,'' J. Acoust. Soc. Am. **110**, 2722(A).

Southall, B. L., Schusterman, R. J., and Kastak, D. (**2000**). ''Masking in three pinnipeds: Underwater, low-frequency critical ratios,'' J. Acoust. Soc. Am. **108**, 1322–1326.

Terhune, J., and Ronald, K. (**1971**). ''The harp seal *Pagophilus groenlandicus* (Erxleben 1777). X. The air audiogram,'' Can. J. Zool. **49**, 285–290.

Terhune, J. M., and Ronald, K. (**1975**). ''Masked hearing thresholds of ringed seals,'' J. Acoust. Soc. Am. **58**, 515–516.

Terhune, J. M. (**1991**). ''Masked and unmasked thresholds of a harbor seal listening in air,'' Can. J. Zool. **69**, 2059–2066.

Turnbull, S. D., and Terhune, J. (**1990**). ''White noise and pure tone masking of pure-tone thresholds of a harbor seal listening in air and under water,'' Can. J. Zool. **68**, 2090–2097.

Ward, W. D. (**1997**). ''Effects of high-intensity sound,'' in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley, New York), pp. 1497–1507.

Effect of coupled oscillations on microbubble behavior

John S. Allen, Dustin E. Kruse, Paul A. Dayton, and Katherine W. Ferrara[a)]
Department of Biomedical Engineering, University of California at Davis, One Shields Avenue, Davis, California 95616-5294

(Received 2 August 2002; revised 20 June 2003; accepted 23 June 2003)

Ultrasound contrast agents are encapsulated microbubbles whose nonlinear acoustic scattering signatures have been the foundation of their use in diagnostic imaging. The coupled oscillations of microbubbles along their lines of center are investigated theoretically using radial equations in the monopole approximation and an energy balance, which is obtained for the system. Coupled microbubble pairs of different initial radii are investigated numerically relative to the normal modes for the linearized system. For microbubble pairs of different size bubbles driven below the mode of the smaller bubble and above the mode of the larger bubble, it is shown that oscillations of the smaller agent are affected substantially more by the coupling than those of the larger one. For separation distances of 10 and 500 microns, a difference of approximately 10 dB occurs in the second harmonic output of a 1.0-micron radius agent coupled with a 2.2-micron radius agent forced at 2.0 MHz and 0.3 MPa. The subharmonic spectral peak is shown to decrease approximately 19 dB for the coupling of 1.5- and 2.2-micron radius agents at 10- and 500-micron distances under the same acoustic forcing conditions. These coupling effects on the radiated pressure and its spectral power are highlighted for contrast agent imaging applications. [DOI: 10.1121/1.1600721]

PACS numbers: 43.80.Qf, 43.25.Yw [FD]

LIST OF SYMBOLS

b	van der Waals hard core radius
d	separation distance
E_o	initial energy of the system
L	Lagrangian
P_o	ambient pressure of the surrounding liquid
$P_g(R)$	pressure of the interior gas of the bubble
$P(R_{ij})$	pressure of bubble i at radius j where j indicates the inner $(j=1)$ or outer $(j=2)$ radius
P_{ijkg}	initial interior gas pressure of bubble i, where j indicates the inner $(j=1)$ or outer $(j=2)$ radius, and the addition of index $k=0$ indicates the initial condition.
r	radial coordinate
R_{ijk}	radius of bubble i, where j indicates the inner $(j=1)$ or outer $(j=2)$ radius, and the addition of index k indicates the initial condition.
$\dot{R}$	bubble wall velocity
$\ddot{R}$	bubble wall acceleration
$T_{i,rr}$	radial stress component, where $i=f$ indicates the stress component for the fluid and $i=s$ indicates the stress component for the shell
T	kinetic energy
U	potential energy
u	fluid velocity
V_i	volume of bubble i
v	volume variable of integration
V	volume of the liquid
μ_i	shear viscosity, $i=f$ indicates the viscosity of the fluid and $i=s$ indicates the viscosity of the shell
κ	polytropic exponent
ρ	density of the surrounding liquid
σ_{ij}	interfacial tension of bubble i, where j indicates the inner $(j=1)$ or outer $(j=2)$ radius
ω_{i0}	linear resonance frequency for bubble i
ω	normal mode frequency

I. INTRODUCTION

Ultrasound contrast agents are encapsulated microbubbles (1–5 microns in radius) with a polymer, protein, or lipid shell. These microbubbles were originally developed for diagnostic imaging purposes to distinguish the acoustic scattering signatures of blood from those of tissue. This capability relies on the agents' nonlinear periodic responses generated at sufficient pressure amplitude by nonlinear oscillations about their equilibrium radii. The scattering signatures from these responses such as harmonics and subharmonics form the foundation for contrast-aided imaging methods.[1]

While these methods, commonly known as harmonic imaging, rely on the agents to remain intact, the destruction and fragmentation of the agents are necessary for other applications. Ultrasound contrast agents have been used in tissue perfusion studies, which monitor the replenishment of agents in vascular regions after the destruction of all detectable agents within the sound field.[1] Such methods may prove to be highly beneficial in investigating and mapping blood flow in capillaries of tumors for the study of angiogenesis.

a)Electronic mail: kwferrara@ucdavis.edu

Also, ultrasound contrast agents have emerged as a potential means for localized drug delivery.[2] These therapeutic agents are designed to release drugs at specific targeted sites. In passive targeting, an external acoustic field is required for the release or activation of the drugs at the specific site.[3,4] Local release often depends on inducing agent fragmentation concurrent with the agent's presence within the tumor's vessels. Moreover, the other method is known as active targeting such that the agents bind to specific sites through ligands that are attached to their surface.[2]

The theory of contrast agent dynamics originates from the fields of physical acoustics and bubble dynamics. The governing equations are similar to those developed earlier to study acoustic cavitation. The theoretical description of the agent's spherical dynamics includes the effect of the shell in a generalized Rayleigh–Plesset formulation. Semiempirical approaches were originally introduced to account for the elastic and dissipative contributions from the shell. More rigorous approaches[5] are gaining favor, as complementary experimental methods to measure shell properties are being proposed and explored. A radial dynamics equation has since become ubiquitous in the contrast agent literature. The equation describes a single agent in an infinite fluid; thus, it is assumed that the agent is located far from boundaries and neighboring agents. Furthermore, this type of equation has been used in numerical simulations of imaging schemes and for both direct and indirect comparisons with experimental results. Features of interest in the radial responses, such as harmonics, may be determined as a function of different acoustic forcing parameters.

However, practical and clinical applications of contrast agents often involve multiple agents in clouds or clusters. Low amplitude, nondestructive scattering from clouds of agents has been used to obtain shell material parameter values through empirical data fitting.[6] The typical theoretical approach used to evaluate multiple agents involves computing the acoustical scattering cross section of a single agent and extending this result to a cloud of agents of the same size. Usually, the distance between the agents is assumed to be sufficiently large such that each of the agents scatters independently. At lower acoustic pressure amplitudes where an agent oscillates harmonically, the corresponding linear acoustic cross sections may be obtained analytically. This approach has been extended to the investigation of the second harmonic responses for a cloud of encapsulated bubbles using a weakly nonlinear perturbation method.[5] The corresponding nonlinear cross sections of secondary harmonic resonances may also be obtained using numerical solutions to a radial dynamics equation.[7] This latter method is outlined in the recent monograph by Angelesen.[8]

These approaches do not account for interactions between the individual agents and possible nonlinear coupling. Moreover, a bubble's dynamics may be altered by the close proximity of neighbors as multiple scattering occurs from the acoustic waves radiated by the nearby bubbles or agents. This close proximity gives rise to a secondary radiation force known as the Bjerknes force after its discoverer V. F. K. Bjerknes.[9] The force arises between two or more agents and depends on the agent size and acoustic forcing parameters, where this force may be either attractive or repulsive. An attractive force can cause agents to translate toward each other such that clusters or clouds are formed. An analytical expression for the Bjerknes force between two agents may be obtained through linear acoustic scattering calculations.[10,11] In this case, the sign of the force may be found analytically as a function of the agent's size relative to the driving frequency. The nonlinear coupling of the oscillations influences the magnitude and direction of the force at higher radial amplitudes such that the linear expressions are no longer valid in this regime.

The investigation of interaction effects in the behavior of bubble clouds and clusters is a subject of ongoing experimental and theoretical study. The theoretical problem has been examined more extensively in terms of linear responses for such applications as oceanic acoustics. A coupled linear oscillator formalism has been developed for two bubbles, and extensions are possible for multiple bubbles. The linear eigenmodes of bubble clouds with equal initial diameter bubbles in spherical configurations have been computed.[12] The fundamental resonance of a uniform bubble cloud has been shown to have a lower frequency than that of a single, isolated agent. Moreover, polydispersed distributions of bubbles can lower the fundamental resonance frequency even further.[13] These theoretical predictions have been confirmed with experimental measurements of acoustic scattering from bubble clouds.[13]

The derivation of the nonlinear equations for multiple bubble dynamics becomes complicated such that approximations are necessary to account for the interactions between the volume and higher-order shape modes of the each of the bubbles. While the experimental work often involves multiple agents in various configurations, theoretical work has concentrated on the interaction between two spherical bubbles along their lines of center due to the complexity of the problem. Usually approximations include only the monopole (breathing mode) and/or dipole (translational) mode. The truncation restricts the motion to spherical oscillations with translation. A notable exception to this has been the numerical study of nonspherical oscillations of two bubbles using a hybrid boundary element method.[14] The free oscillations were examined along with forced oscillations produced by an impulse response. The complexity of the numerical method limited the number of oscillation cycles that could be studied.

In the investigation of forced oscillations, research has concentrated on spherical assumptions. Using the virial theorem for a potential flow, researchers have derived a set of equations for nonlinear oscillations and the translation of two bubbles about their lines of center.[15] The equations obtained simplify to two coupled radial equations in the limiting case where only the monopole interaction is significant. A separate derivation of these monopole interaction equations and subsequent extensive numerical parameter study of the Bjerknes force has been reported more recently.[16] The dynamics and oscillations of the Bjerknes force arising from millimeter-size bubbles forced at low-kHz frequencies were investigated with respect to the formation of acoustic streamers. The two bubble equations for a monopole interaction

have also been derived by a Lagrangian method.[17] These equations were examined with respect to their normal modes in the limit of small-amplitude oscillations. The terminology warrants some explanation to avoid confusion. The normal modes here refer to the eigenmodes arising from a system with two (or more) degrees of freedom. The particles for each normal mode oscillate with the same frequency. This should be differentiated from terms which describe the individual resonances of a single bubble in terms of its deformation, i.e., the volume and shape modes.

These theoretical developments regarding coupled interactions and a generalized Bjerknes force for unencapsulated gas bubbles have not been extensively investigated with respect to ultrasound contrast agent applications. Coupled agents will influence each other's dynamics and should in turn modify the net radiated acoustic pressure. Potentially different acoustic radiation characteristics could provide for improved detection schemes since tissue harmonics generated by nonlinear propagation are not expected to have such coupling properties. Emerging potential drug delivery applications would benefit from an understanding of acoustic signatures of a group of interacting agents. Efficacy of active targeting of drugs might be monitored by scattering signatures unique to agents within close proximity to each other. The effect of bubble clouds on tissue and cells has yet to be fully characterized, and the potential beneficial or deleterious effects on drug delivery warrant investigation. Furthermore, there exists a disparity between experimental observations of multiple scattering effects, such as the Bjerknes force, and the corresponding theory used in ultrasound contrast agent studies. As mentioned previously, the standard treatment of multiple contrast agents neglects any coupling in the radial oscillations. The study of the oscillation of two coupled agents provides a preliminary foundation to address some of these issues. The paper is organized as follows: the theory for two bubbles with shells (contrast agents), methods, results, discussion, and conclusions.

II. THEORY

This work will focus on the monopole interaction between two contrast agents or microbubbles. The governing equations follow directly from those developed for unencapsulated gas bubbles. With the assumptions outlined below, the nature of the nonlinear interaction terms is not affected by the inclusion of a shell. The optimal description of the contrast agent shell has yet to be determined, as the specific material parameter values for the various shell types remain unresolved. Nevertheless, the inclusion of the elastic and viscous contributions from a previously developed formulation[5] may be incorporated into the equations of the coupled agents for completeness. The inclusion of this formulation is feasible with some recent theoretical extensions that allow for the study of shells of arbitrary thickness.[4]

The coupled equations for two spherical bubbles may be derived from considerations of the scattered pressure felt by a neighboring agent.[16] For a numerical study of coupled oscillations and the corresponding Bjerknes force, this derivation is carried out in the incompressible limit, and first-order compressibility corrections are subsequently added in an *ad hoc* manner in an attempt to more accurately model the large amplitude oscillations. This *ad hoc* introduction of liquid compressibility does not account for a variable interaction time that may exist between bubbles due to the introduction of a finite speed of sound.[16] Further study is needed to determine the validity of this approximation for the parameter range of this paper; however, this is beyond the scope of this study.

An alternative derivation of the two-bubble equations from a Lagrangian formulation was introduced in an earlier work.[17] The Rayleigh–Plesset equation for a single bubble may be obtained using this formulation such that the inclusion of another bubble coupled to the first is a logical extension. The Lagrangian consists of the potential and kinetic energy of the system and the equations of motion are obtained from this Lagrangian. The addition of another bubble in close proximity modifies the kinetic energy of the system through the scattering interaction that occurs between the two agents. The equations allow for small distances between the bubbles assuming they oscillate spherically. Additional equations are required to incorporate coupled translational effects. For this study, bubbles at a fixed separation distance will be examined. The higher-order terms that correspond to translation are not likely to directly contribute to radiated pressure in the far field. Such a contribution will affect the dynamics if the radial oscillations are significantly coupled to the translation. In the derivation, the system is assumed to be conservative from the Lagrangian viewpoint. The viscous effects of the surrounding fluid and the shell do not need to be introduced into the Lagrangian. These are included through the normal stress balance at the bubble wall such that they may be introduced into the governing equations upon obtaining the differential equations. We follow this approach and highlight the inclusion of a contrast agent shell into the formulation.

The velocity field of the system can be expressed as the summed velocity of the two bubbles arranged to interact along their lines of center

$$u_1+u_2=\frac{\dot{R}_1R_1^2}{r_1^2}+\frac{\dot{R}_2R_2^2}{r_2^2}. \tag{1}$$

The kinetic energy T of the system is expressed in terms of velocity of the system as an integral over the entire volume (V) of the liquid

$$T=\frac{\rho}{2}\int_V(u_1+u_2)^2dV. \tag{2}$$

The density of the surrounding liquid is denoted by ρ. The interaction due to the cross terms in (2) may be truncated on the order of the monopole terms assuming the separation distance is greater than the sum of the radii. With the substitution of (1) into (2) and completing the truncated integral, the kinetic energy is expressed as follows:

$$T\approx 2\pi\rho\left(R_1^3\dot{R}_1^2+R_2^3\dot{R}_2^2+\frac{2R_1^2R_2^2}{d}\dot{R}_1\dot{R}_2\right). \tag{3}$$

Here, d is the distance between the two bubbles along their lines of center. The potential energy, U, is determined by the difference of internal gas pressure and the surrounding am-

bient liquid pressure for each of the bubbles. The quantity is defined as the sum of the integrals computed over the volume of each of the bubbles

$$U = -\int_{V_1} (P(v_1) - P_o) dv_1 - \int_{V_2} (P(v_2) - P_o) dv_2. \quad (4)$$

The Lagrangian is expressed as the difference of the kinetic and potential energies

$$L = T - U. \quad (5)$$

For this case, (3) and (4) determine the Lagrangian

$$L = 2\pi\rho\left(\dot{R}_1^2 R_1^3 + \dot{R}_2^2 R_2^3 + \frac{2}{d} R_1^2 R_2^2 \dot{R}_1 \dot{R}_2\right) + \int_{V_1} (P(v_1) - P_o) dv_1 + \int_{V_1} (P(v_2) - P_o) dv_1. \quad (6)$$

The generalized coordinates for the Lagrangian are the radius and radial velocities of each of the bubbles. The governing equations of motion are obtained from the standard procedure

$$\frac{\partial L}{\partial R_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{R}_i} = 0; \quad i = 1,2. \quad (7)$$

The equations of motion are obtained for the coupled radial oscillations of the two bubbles

$$\begin{aligned} R_1\ddot{R}_1 + \frac{3}{2}\dot{R}_1^2 + \frac{R_2}{d}(R_2\ddot{R}_2 + 2\dot{R}_2^2) &= \frac{1}{\rho}(P(R_1) - P_o), \\ R_2\ddot{R}_2 + \frac{3}{2}\dot{R}_2^2 + \frac{R_1}{d}(R_1\ddot{R}_1 + 2\dot{R}_1^2) &= \frac{1}{\rho}(P(R_2) - P_o). \end{aligned} \quad (8)$$

These reduce to a set of uncoupled Rayleigh bubble dynamics equations in the limit that the distance, d, between the bubbles becomes infinitely large. The coupled interaction provides an additional inertial contribution through the terms inversely proportional to the distance. Following the convention to close the bubble dynamics equations, the pressure in the bubble interior needs to be specified. The interior gas pressure is related to the pressure at the bubble wall through the normal stress balance. The simplest assumption for the interior pressure is that an ideal gas obeys a polytropic gas exponent relationship. The approximation has known limitations in accounting for the thermal damping due to the heating of gas; nevertheless, it continues to be one of the most commonly used.[7] Modified versions of this expression with the van der Waals gas terms have been introduced in order to more realistically accommodate gas behavior during rapid bubble collapses.[7] The van der Waals gas pressure may be expressed as follows, where b represents the hard core radius:

$$P_g(R) = P_{go}\left(\frac{R_o^3 - b^3}{R^3 - b^3}\right)^{\kappa}. \quad (9)$$

The inclusion of the contribution of a contrast agent shell into a single bubble dynamics equation extends the standard derivation of the Rayleigh–Plesset equation. This follows from an integration of momentum equation over the shell thickness, with the shell material contributing to the radial stress balance at the interfaces.[5] In following this derivation, any density difference between the shell and the surrounding liquid contributes to additional inertial terms. However, the coupled two-bubble equations described above are derived from a different approach, that of using a Lagrangian. To properly account for these additional inertial terms due to the shell using this method, the expression for the kinetic energy (2) would need to be modified to reflect the density difference between the shell and the outer surrounding fluid. However, if the density of the shell is the same or very close to that of the surrounding fluid, these terms become negligible. Since the density properties of the lipids and other materials used for contrast agents are either not well known or estimated to be close to that of water, we neglect these terms. Thus, in this simplified case, the shell contribution may be introduced into the governing equation through the radial stress balances at the inner and outer boundaries. The interior gas balances the inner radius of the bubble such that the radii in (8) should be taken as the inner radii. We introduce the following notation to specify the inner and outer radii of each of the agents. Let the first subscript denote the bubble and the second the inner or outer radii of the shell of the respective bubble, denoted by 1 and 2, respectively. Using this notation, the pressure at each of the bubble walls may be expressed as follows:

$$\begin{aligned} P(R_{12}) &= P_g(R_{11}) - \frac{2\sigma_{11}}{R_{11}} - \frac{2\sigma_{12}}{R_{12}} + 3\int_{R_{11}}^{R_{12}} \frac{T_{s,rr}}{r} dr \\ &\quad + 3\int_{R_{12}}^{\infty} \frac{T_{f,rr}}{r} dr, \\ P(R_{22}) &= P_g(R_{21}) - \frac{2\sigma_{21}}{R_{21}} - \frac{2\sigma_{22}}{R_{22}} + 3\int_{R_{21}}^{R_{22}} \frac{T_{s,rr}}{r} dr \\ &\quad + 3\int_{R_{22}}^{\infty} \frac{T_{f,rr}}{r} dr. \end{aligned} \quad (10)$$

The constitutive equation for the shell and surrounding fluid determines the radial stress contributions. The constitutive equation is assumed to be traceless such that for a spherically symmetric bubble it may be expressed only in terms of the radial component.[5] The subscript "s" refers to the shell and "f" denotes the surrounding fluid, which contributes to the surface traction at the outer shell boundary. Surface tension occurs at the inner and outer boundaries of the shell material; however, the values corresponding to the shell interior have not been established. For a relatively thin shell, these terms act almost identically such that the surface tension may be represented by taking only the outer value with respect to the inner radius.[4] Using the constitutive equation for a Newtonian fluid, the radial stress component may be expressed as follows:

$$T_{rr} = -2\mu\frac{\partial u}{\partial r}. \quad (11)$$

The surface traction contribution due to the surrounding fluid may be expressed as follows for each of the respective agents:

$$3\int_{R_{12}}^{\infty}\frac{T_{f,rr}}{r}=-4\mu_f\frac{R_{11}^2\dot{R}_{11}}{R_{12}^3},\quad 3\int_{R_{21}}^{\infty}\frac{T_{f,rr}}{r}=-4\mu_f\frac{R_{21}^2\dot{R}_{21}}{R_{22}^3}. \tag{12}$$

The constitutive properties of the shell are not well understood, and the shell is usually modeled as an elastic-viscous material such that the radial stress tensor may be divided into separate elastic and viscous contributions. In this work, we approximate the lipid shell as a viscous Newtonian fluid. The constitutive equation for a viscous fluid given by (11) may be integrated over the shell region and expressed as follows:

$$\begin{aligned}3\int_{R_{11}}^{R_{12}}\frac{T_{s,rr}}{r}dr&=-4\mu_s\dot{R}_{11}\left[\frac{R_{12}^3-R_{11}^3}{R_{12}^3R_{11}}\right],\\ 3\int_{R_{21}}^{R_{22}}\frac{T_{s,rr}}{r}dr&=-4\mu_s\dot{R}_{21}\left[\frac{R_{22}^3-R_{21}^3}{R_{22}^3R_{21}}\right].\end{aligned} \tag{13}$$

While the viscous component follows directly from integration of constitutive equation (10), the elastic component has previously been included in the contrast agent model in an approximate manner.[5] The neglect of elastic component and exact inclusion of the viscous component allows for a formulation of the energy balance as outlined below.

A recent development has been the use of an expression to close the equations based on the fact that motion of the incompressible system is isochoric.[4] This conservation of volume allows the outer radius to be expressed in terms of the inner radius. The first bubble is denoted by a first subscript of "1," the second subscript denotes the inner or outer radius, and the equilibrium inner and outer radius are denoted by "0," a third subscript. The resulting expressions are given by (14) and (15)

$$R_{12}=(R_{11}^3+R_{120}^3-R_{110}^3)^{1/3}, \tag{14}$$

$$R_{22}=(R_{21}^3+R_{220}^3-R_{210}^3)^{1/3}. \tag{15}$$

The expressions (14), (15) may be substituted into (12) and (13) to express the integrals entirely in terms of the inner radius.

The system exchanges energy back and forth between the two bubbles since they are coupled through the monopole interaction term. Though the Lagrangian used is conservative, the system is not due to dissipation from the viscous surrounding fluid and shell. An energy balance can be expressed in terms of the sum of the potential and kinetic energies and integrating over the dissipation and work done by acoustic forcing. With the use of a polytropic exponent relationship for the gas, it is easy to express the potential energy of the interior gas. The kinetic energy corresponds to that of each of the individual bubbles, and in addition, the kinetic energy of interaction terms. The dissipation comes from two contributions, that of the surrounding fluid and that of the shell, since thermal and radiation losses have not been included at this point. The energy balance may be obtained from the governing equations by multiplying them by a surface area term of $4\pi\dot{R}_iR_i^2$. The energy balance may be expressed exactly using (12) and (13) and may be expressed solely in terms of the inner radius with the use of (14) and (15). The energy balance can be expressed as follows (with one surface tension parameter):

$$\begin{aligned}&2\pi\rho\left(\dot{R}_{11}^2R_{11}^3+\dot{R}_{21}^2R_{21}^3+\frac{2}{d}R_{11}^2R_{21}^2\dot{R}_{11}\dot{R}_{21}\right)+\frac{4\pi}{3}R_{11}^3\left(P_0+\frac{P_{110g}(1-b)^{\kappa}}{(3-3\kappa)}\left(\frac{R_{110}^3-b^3}{R_{11}^3-b^3}\right)^{\kappa}+\frac{3\sigma}{R_{11}}\right)\\ &\quad+\frac{4\pi}{3}R_{21}^3\left(P_0+\frac{P_{210g}(1-b)^{\kappa}}{(3-3\kappa)}\left(\frac{R_{210}^3-b^3}{R_{21}^3-b^3}\right)^{\kappa}+\frac{3\sigma}{R_{21}}\right)\\ &\quad=-4\pi\int_0^t\dot{R}_{11}R_{11}^2p(t)dt-4\pi\int_0^t\dot{R}_{21}R_{21}^2p(t)dt-4\pi\mu_f\int_0^t\frac{\dot{R}_{11}^2R_{11}^4}{R_{12}^3}dt-4\pi\mu_f\int_0^t\frac{\dot{R}_{21}^2R_{21}^4}{R_{22}^3}dt\\ &\quad-4\pi\mu_s\int_0^t\dot{R}_{11}^3R_{11}\left[\frac{R_{12}^3-R_{11}^3}{R_{12}^3}\right]dt-4\pi\mu_s\int_0^t\dot{R}_{21}^3R_{21}\left[\frac{R_{22}^3-R_{21}^3}{R_{22}^3}\right]dt+E_0.\end{aligned} \tag{16}$$

The first term on the left-hand side represents the kinetic energy of the system. The potential energies of bubbles one and two are represented by the next two terms on the left-hand side. The work done by forcing is given by the first two terms on the right-hand side, where $p(t)$ is the acoustic pressure waveform. The third and fourth terms on the right-hand side are the viscous dissipation due to the surrounding fluid, and the fifth and sixth terms are the viscous dissipation due to the shell. The last term is the initial energy of the system. The energy balance provides a check to monitor the accuracy of the calculation as it is computed simultaneously along with the radial equations. The transfer of energy and dissipation of the system can be investigated by examining the oscillations of the corresponding components for a physical understanding of the system.

The normal modes of the system may be obtained from

the solution of the linearized version of the coupled equations given by (8).[17] The normal modes may be expressed as the roots of the following:

$$(\omega_{10}^2-\omega^2)(\omega_{20}^2-\omega^2)-\frac{\omega^4 R_{110}R_{210}}{d^2}=0. \tag{17}$$

The terms ω_{10}, ω_{20} represent the natural uncoupled linear resonance frequencies of bubbles one and two. The expression simplifies when the bubbles are of equal size. The normal modes provide a physical description for two bubbles much the same way the linear resonance frequency does for a single, isolated bubble, though only for specific initial conditions do the bubbles oscillate at the same frequency. The general solution to the system's linear oscillations is usually composed of a superposition of the two modes.

The liquid compressibility corrections have not been included in the coupled equations. The rigorous inclusion of compressibility corrections for the liquid provides additional complications such that the structure of the equations may change due to time delays.[16] Thus, the equations of the coupled bubbles become differential delay equations rather than ordinary differential equations. Since the coupled bubble dynamics equations have not been examined before in the context of contrast agent applications, nor compared directly with experimental results, we will restrict ourselves to the incompressible case. For contrast agent imaging applications, a driving pulse is applied and the effect of delays may be minimal if they are not given sufficient time to accumulate. Previous studies of the dynamics of a single bubble indicate for radial excursions less than approximately 2.2–2.5 times the equilibrium radius,[18] the incompressibility assumptions are suitable. This restriction is used in selecting parameters for the theoretical studies and results presented. To examine some higher pressure amplitude forcing cases, *ad hoc* compressibility assumptions were added to the equations. The results should be viewed with that caveat.

Independent measurements of contrast agent shell properties have yet to be obtained, so empirical fits to experiment data give the best estimate for these values. Previously, statistical fits were used to determine a product of the shell thickness and viscosity.[19] With the recent description of shells of finite thickness, both the thickness and shell viscosity are independent parameters in the problem.[4] Elasticity has not been included in the present work as its variation with shell thickness for lipids has not been established, and previous work has suggested that a lipid shell may be reasonably approximated as a viscous fluid.[19]

III. METHODS

For the calculations, the values for the surface tension, van der Waals gas constant, and other parameters have been set to values previously established for lipid-shelled contrast agents with a high molecular weight gas interior.[20] Acoustic and bubble parameters have been chosen for the numerical calculations to investigate bubbles of initial radii 1.0 to 3.0 microns. For these studies, the viscosity value of the shell was selected to be within the range of previous empirical fits[19] and a 2-nm thickness was assumed. To observe novel nonlinear interactions and not significantly exceed the incompressibility conditions, a value of 3.0 poise was chosen. An 8.0-cycle sinusoidal pulse was used along with a pressure amplitude of 0.3–0.5 MPa and a frequency of 2.0 MHz. Shorter cycle pulses were examined; however, a sufficiently longer time was needed for the different coupling components to become visible in the spectrum.

As a first-order approximation, the radiated pressure can be treated as the sum of the pressure contributions from the individual monopole potentials of each of the two agents. The pressure expression may be obtained from the respective velocity potentials of the two agents.[7] Within the monopole approximation for the coupled equations, higher-order interactions influence the net pressure radiated by the system; however, the above assumption is reasonable for distances far from the bubble. Of particular interest for contrast agent applications is the spectrum computed from this radiated pressure, and its associated power. Equations (8), with the inclusion of (9) and (13)–(15), are solved numerically with the software package MATLAB. The energy balance solution (16) is obtained from the solution of the radial equation and the numerical computation of the associated integrals.

The dynamics of two bubbles of the same size has been the subject of the most research. This has been examined in the framework of linear acoustics[21] and with some extensions to examine the role of nonlinearities in the coupling.[22] In these works, behavior near the normal modes has been emphasized since (17) simplifies in this case, facilitating physical interpretation of the behavior. Previous numerical work on coupled nonlinear equations examined forcing below both modes.[16] The dynamics and the pressure radiated from coupled different-sized bubbles are of interest to contrast agent applications. While the dynamics of a single bubble is described relative to its linear resonance frequency, the two-bubble situation is more complicated because of the combination of mode frequencies. We examine in detail the forcing between modes, as this has received less attention and the physical characteristics may be investigated with respect to the potential physical limitations of the equations.

IV. RESULTS

Theoretical cases are investigated for agents of differing initial radius forced below one mode and above both modes.

A. Forcing between and above the normal modes

Figure 1 provides a plot of the frequencies (nonangular) corresponding to the normal modes of the system for a pair of agents with one initial radius fixed at 2.2 microns and a variable radius, and separation distances of 10 and 500 microns. The mode frequencies for the two separation distances have a similar shape, with the effect of separation distance observed most clearly when the agents are the same or similar size. When both agents are the same size (2.2 microns), the mode frequencies converge for the 500-micron separation distance, while there is a difference in mode frequency for the 10-micron separation distance. Note that the first mode is relatively flat, with a value of 1.5 MHz at an initial radius of 0.5 microns, decreasing to 0.6 MHz at 5.0 microns. The second mode decreases rapidly from a frequency of 9

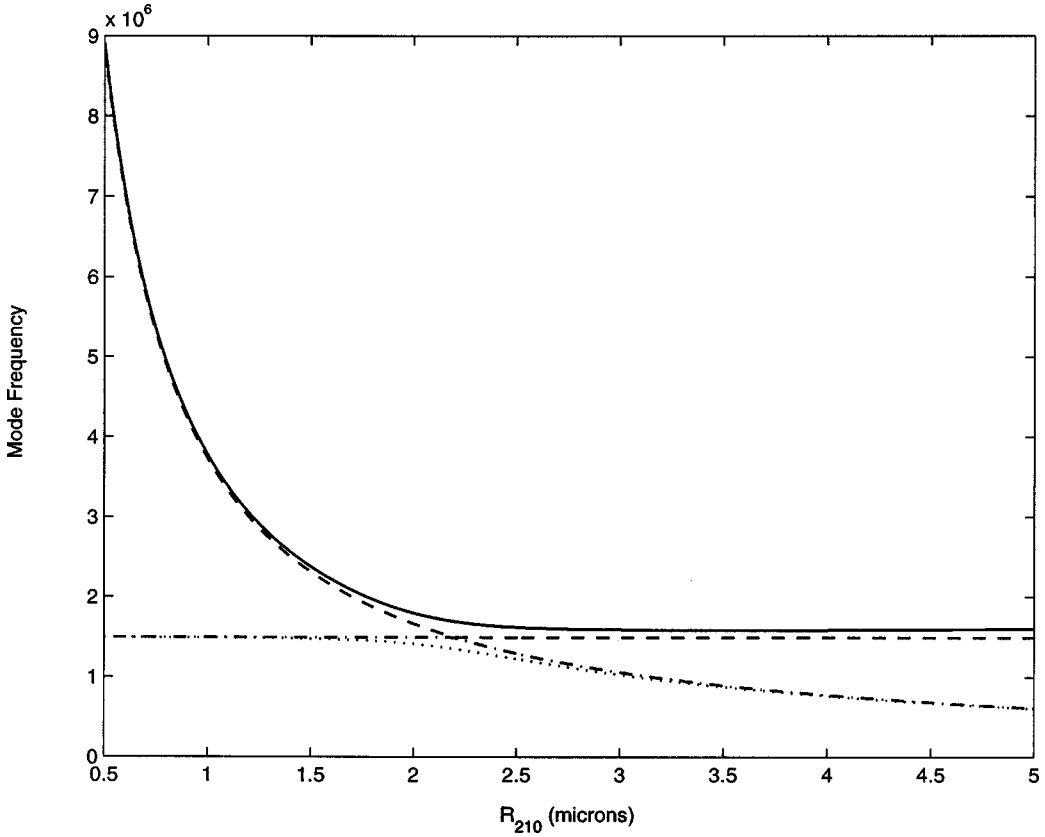


FIG. 1. The normal modes of two agents are plotted for separation distances of 10 and 500 microns for one radius fixed at 2.2 micron and the other varied from 0.5 to 4.0 microns. These are denoted as the first mode at 10 microns (dotted line), the second mode at 10 microns (solid line), and the first mode at 500 microns (dash-dotted line) and the second mode at 500 microns (dashed line).

MHz to 2.0 MHz over the range of initial radii from 0.5 to 2.0 microns, respectively. A forcing frequency of 2.0 MHz is above the first mode and below that corresponding to the second mode for the 1.0- and 2.2-micron agents. The distance to the upper mode is less for agents of 1.5 and 2.2 microns forced at 2.0 MHz. Figure 2 shows a 0.3-MPa, 2.0-MHz, eight-cycle forcing pulse used for the study. The higher pressure amplitude pulses used in the study have a similar shape.

Figure 3 shows the radial oscillations of 1.0- and 2.2-micron agents at separation distances of 10 and 500 microns. Differences are observed in the radial oscillations of the 1.0-micron initial radius agent at these separation distances, where this case is shown in Fig. 3(a). With the 500-micron separation, the radial amplitude excursions increase (compared with the 10-micron separation) and are nearly uniform in amplitude over time. In contrast, for the 10-micron separation distance, the radial excursions have an irregular shape with small secondary oscillations. The corresponding negative wall velocities over the driving pulse are approximately 10 to 44 m/s at the 10-micron separation distance and 110 to 150 m/s at the 500-micron separation distance. Also, the ring-down after 5.0 microseconds requires a longer time interval at the 10-micron separation distance. Figure 3(b) provides the radial oscillations of the 2.2-micron radius agent (while coupled to the 1.0-micron agent) at separation distances of 10 and 500 microns. The two curves are similar in shape except for a few subtle differences. Amplitude excursions near 3.0 and 4.0 microseconds are diminished at the 500-micron separation distance, compared with the 10-micron distance.

The changes in the radial oscillations of the 2.2-micron radius agent, as a function of separation distance, result in little change in the radiated pressure and its corresponding spectra, which therefore are not shown. However, substantial differences occur in the spectra of the 1.0-micron radius agent, as a function of separation distance. The spectra of echoes from the 1.0-micron agent are shown in Fig. 4(a) for separation distances of 10, 20, and 500 microns. At the 10-micron distance, the 2/1 and 3/1 harmonics of the fundamental are visible near 4.0 and 6.0 MHz. At this separation distance, additional peaks occur near 3.0 and 5.0 MHz. These additional peaks significantly diminish in height for the greater separation distances of 20 and 500 microns. Figure 4(b) shows the total pressure spectrum from both agents. Notice a substantial difference ($\sim$10 dB) for the distances of 10 and 500 microns in the peaks heights near 4.0 and 6.0 MHz.

Figure 5(a) shows the radius time curve of the 2.2-micron agent at 10- and 500-micron separation distances coupled with a 1.5-micron radius agent. The choice of a 1.5-micron radius is made in order to evaluate the effect of the smaller bubble when it is driven near the upper normal mode. The oscillation of the 2.2-micron agent increases to a greater extent at the 500-micron distances as the excursion amplitude becomes more irregular. Figure 5(b) shows the

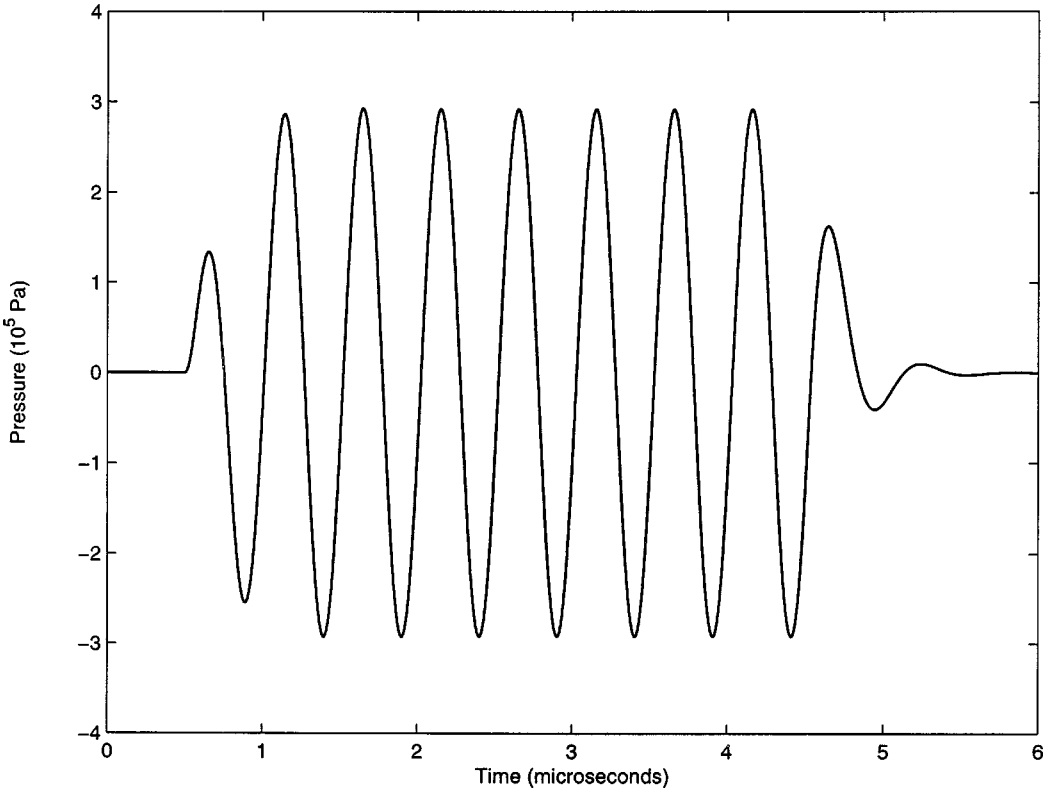


FIG. 2. The eight-cycle pulse of a pressure amplitude of 0.3 MPa and a center frequency of 2.0 MHz used in the theoretical calculations is shown.

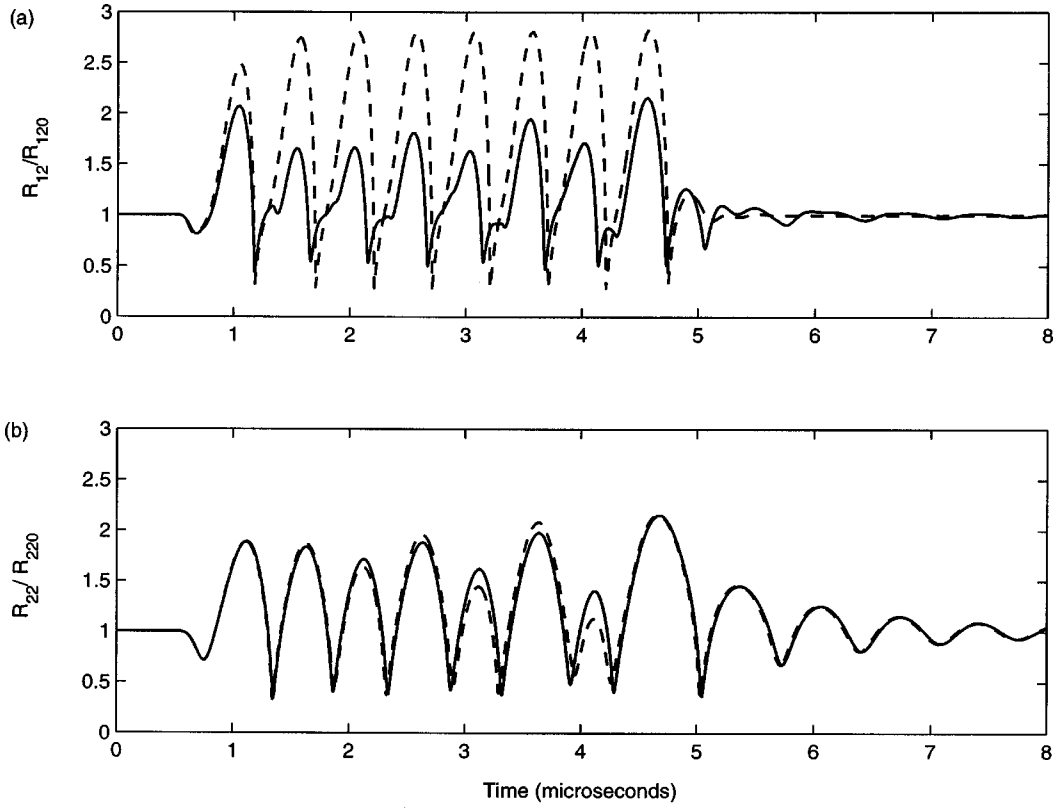


FIG. 3. (a) The radial oscillations of the 1.0-micron radius agent at 10 microns (solid line) and 500 microns (dashed line) separation distances are shown. Notice the oscillations are smaller in amplitude and of irregular shape at 10 microns compared to 50 microns. (b) The radial oscillations of a 2.2-micron initial radius agent coupled with a 1.0-micron agent at separation distances of 10 microns (solid line) and 500 microns (dashed line) are shown. Only subtle differences are visible in the amplitude excursions at these two distances.

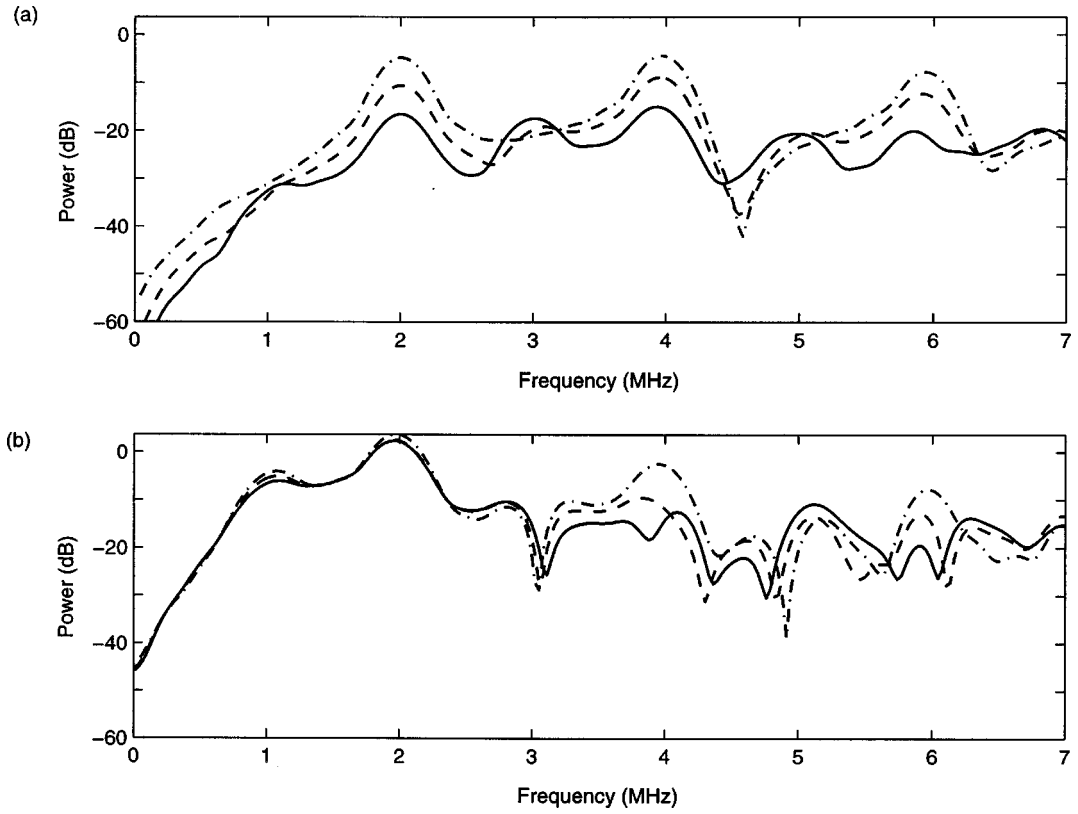


FIG. 4. (a) The spectra of the radiated pressure of the 1.0-micron agent, coupled to the 2.2-micron agent, at the separation distances of 10 microns (solid line), 20 microns (dashed), and 500 microns (dash dot line) are shown. (b) The total spectra of the radiated pressure from the 1.0-micron and 2.2-micron agents coupled together, at the separation distances of 10 microns (solid line), 20 microns (dashed), and 500 microns (dash dot line) are shown. Notice the difference in the peak height at 4.0 and 6.0 MHz as a function of distance.

corresponding radiated pressure spectrum from this agent at 10- and 500-micron distances. The spectral peak near 1 MHz occurring at 10-micron separation distance is attenuated approximately 19 dB at the 500-micron distance.

The integrated power within frequency bands has been used previously to quantify spectral characteristics of harmonic echoes, and is applied here to summarize changes in echo characteristics with varied initial radius and separation distance.[7] In the following two figures, the integrated power within 500-kHz bands and the integrated power over a 5-MHz bandwidth are shown. Figure 6(a) shows the integrated echo power for one of a pair of coupled agents, where one agent has an initial radius of 2.2 microns and the second (the agent generating the echo) radius is varied from 1.0 to 5.0 microns. The separation distance is 500 microns and the acoustic forcing pulse is shown in Fig. 2. Figure 6(b) is a similar plot with a separation distance of 10 microns. The most significant changes between the two plots occur for a small initial radius (i.e., a coupling between the 2.2-micron larger radius and a smaller radius). For initial radii below 2.0 microns, the total integrated power, and the integrated power near the fundamental (2 MHz) and first harmonic (4 MHz) are approximately 10 dB lower for a 10-micron compared with a 500-micron separation distance, indicating energy transfer from these components.

The total power (below 5.0 MHz) and the integrated power from 1.75 to 2.25 MHz follow similar shapes within each figure, and differences between the figures are consistent. In Fig. 6(b), a notable increase in echo power is evident for a 3.0-micron radius agent, compared with the 1.5-micron radius, over frequency ranges 2.75–3.25 MHz and 0.75–1.25 MHz. In addition, a dip is observed in the 3.75–4.25-MHz bin for an initial radius of 3.5 microns at 10-micron separation distance.

Figures 7(a)–(b) show the integrated power for the same parameters as Figs. 6(a)–(b), except at a higher pressure amplitude of 0.5 MPa. The compressibility corrections included for these results produced minor changes in the figures. The power of the frequency bin centered at 1.0 MHz decreases approximately 10 dB at the 10-micron separation distance for agent sizes below 1.5 microns compared with the 500-micron distance. At the 500-micron separation distance, the 4.0-MHz bin decreases for radii from 1.5 to 2.5 microns. In contrast, the bins centered at 1.0 and 2.0 MHz rise sharply over this same interval.

V. DISCUSSION

As noted in the Introduction, it has been assumed previously that multiple contrast agents interact and scatter independently.[8] We extend the investigation to agents of differing initial radius in the nonlinear regime.

A. Forcing between and above the normal modes

An important point to note from Fig. 1 is that when the agents are coupled the linear dynamics is usually governed by the superposition of the two normal modes. For nonlinear coupled oscillations the situation becomes more compli-

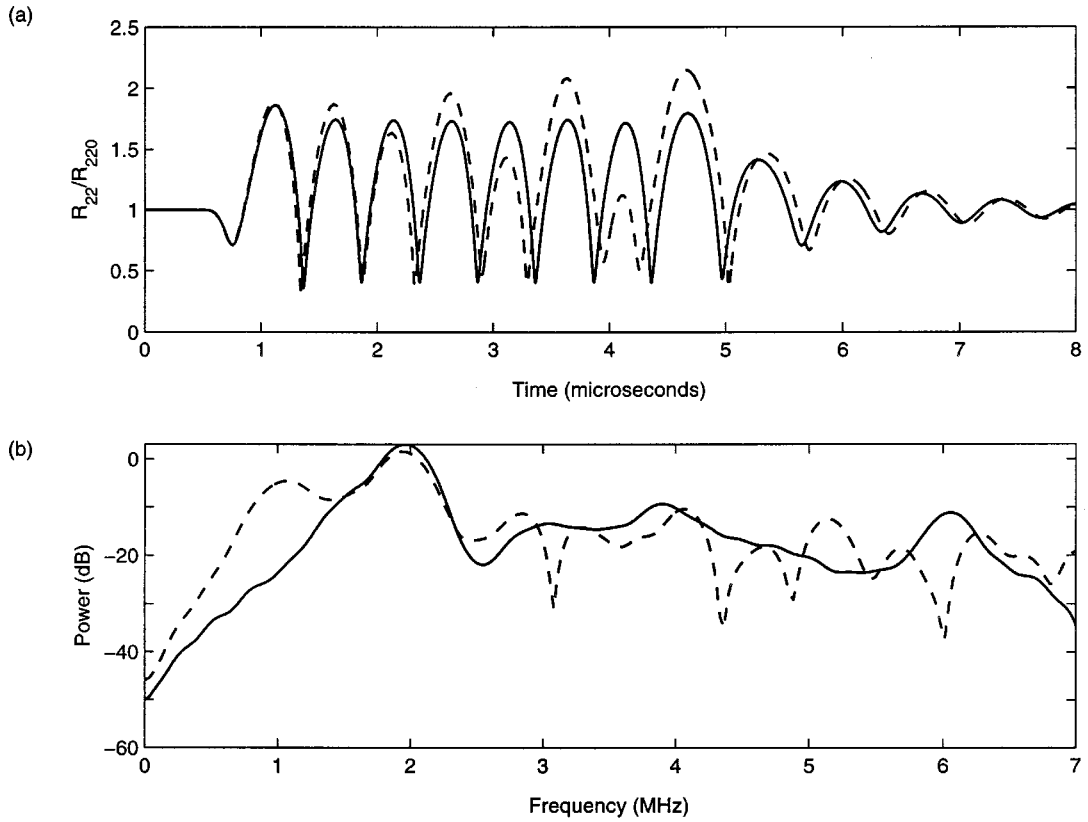


FIG. 5. (a) The radial oscillations of a 2.2-micron initial radius agent (coupled with a 1.5-micron agent) at separation distances of 10 microns (solid line) and 500 microns (dashed line). Note that the radial oscillations are less uniform at the 500-micron distance are shown. (b) The spectra of the radiated pressure of the 2.2-micron agent (coupled to a 1.5-micron agent) at the separation distances of 10 (solid line) and 500 (dashed line) microns are shown. The subharmonic peak at 1.0 MHz diminishes at the closer separation distance of 10 microns.

cated; however, the driving frequency value relative to the modes offers some guidelines. For large separation distances, as seen from (17), the agents act as uncoupled oscillators.

The echo power spectrum was considered for one of a pair of coupled agents in which the pair was forced above one mode and below the other mode. This analysis reveals significant changes as a function of agent size and coupling distance. The oscillation of the smaller (1.0-micron) agent is substantially changed due to coupling. The radial oscillation of the larger (2.2-micron) agent changes less as a function of separation distance than the 1.0-micron radius agent, as seen in Figs. 3(a) and (b). The coupling that is apparent in Fig. 3(a) occurs through a term that represents the inertia of the liquid [third term on the left-hand side of Eq. (8)]. The radial oscillations of the 1.0-micron agent demonstrate small-scale growth and collapse features, which reflect the inertial nature of the coupling terms. These are reminiscent of a single agent driven above its linear resonance.

Moreover, the spectrum of the 1.0-micron agent in Fig. 4(a) reveals some interesting characteristics not readily apparent from the radius versus time curves. A spectral peak is prominent at 3.0 MHz for the case of the 10-micron separation distance, and this diminishes with increasing separation distance. The coupling contributes to fractional harmonics (3/2, 5/2) corresponding to the peaks at 3.0 and 5.0 MHz at the 10-micron separation distance. Energy moves into these harmonic responses, resulting in the diminished second (4.0-MHz) and third harmonic (6.0-MHz) peaks in the total spectrum seen in Fig. 4(b). This result suggests the second harmonic response is sensitive to the separation distance and the coupled agents' size differential, which previously has not been accounted for in theoretical and experimental investigations of contrast agents.

The 2.2-micron agent's radial dynamic changes when it is coupled with a 1.5-micron agent while keeping the other acoustic parameters fixed, as seen in Fig. 5(a). This is the key difference between this and the previous case with the 1.0-micron agent. The 1-MHz amplitude modulation seen in Fig. 5(a) for the 500-micron distance corresponds with the subharmonic response seen in Fig. 5(b). The 1.5-micron agent is forced close to the upper normal mode, and thus this component of the system is easier to drive so it dominates the coupling. The 2.2-micron agent behaves more like a free bubble driven above resonance at the larger separation distance. A subharmonic component develops but the coupling eliminates this response at the 10-micron distance.

Key points are summarized in Figs. 6(a)–(b) with respect to the coupling as a function of the distance and agent size. At the 500-micron separation distance, the agents' oscillatory responses can be viewed as being essentially uncoupled. In this case, the normal modes are close to the natural frequencies of two uncoupled agents. Since the nonlinear harmonic responses of a single agent are well studied, it is straightforward to discern known characteristics from this plot. A comparison of the 500-micron separation distance case with that for a 10-micron separation distance illustrates

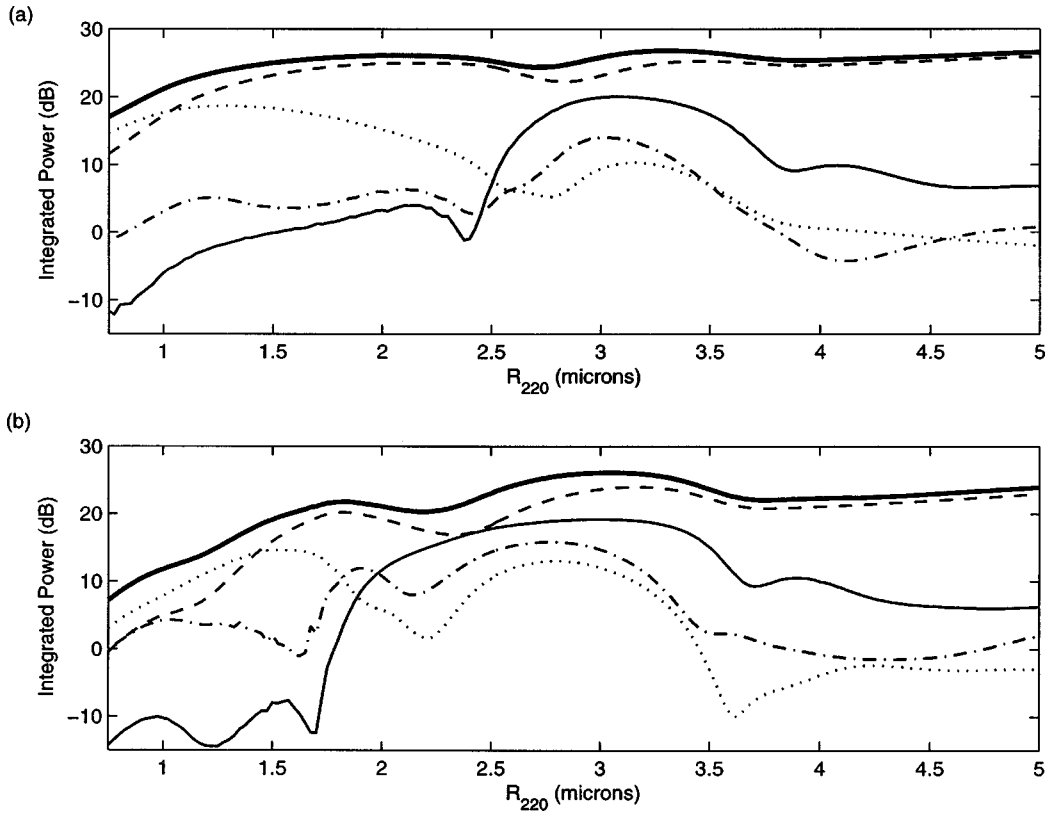


FIG. 6. (a) At a separation distance of 500 microns, integrated power is plotted over 500-kHz bins for a 2.2-micron agent coupled with a second agent of initial radius from 1 to 5 microns. The frequency bins are denoted as below 5 MHz (dark solid line), centered at 1.0 MHz between 0.75 and 1.25 MHz (solid line), centered at 2.0 MHz between 1.75 and 2.25 MHz (dashed line), centered at 3.0 MHz between 2.75 and 3.25 MHz (dash dot line), centered at 4.0 MHz between 3.75 and 4.25 MHz (dotted line). (b) At a separation distance of 10 microns, integrated power is plotted over 500-kHz bins for a 2.2-micron agent coupled with a second agent of initial radius from 1 to 5 microns. The frequency bins are denoted as below 5 MHz (dark solid line), centered at 1.0 MHz between 0.75 and 1.25 MHz (solid line), centered at 2.0 MHz between 1.75 and 2.25 MHz (dashed line), centered at 3.0 MHz between 2.75 and 3.25 MHz (dash dot line), centered at 4.0 MHz between 3.75 and 4.25 MHz (dotted line).

and summarizes the effects of coupling. The effect of coupling a larger agent with a smaller one becomes apparent for initial radii less than approximately 2.0 microns. Coupling with an agent with an initial radius less than 2.0 microns results in a 5–10-dB decrease in echo power at a 10-micron separation distance, compared with 500 microns.

In Fig. 6, the 1.5-micron agent's linear resonance frequency is near the driving frequency of the pulse (2.0 MHz). Its fundamental resonance response and harmonics are evident at the 500-micron separation distance. For a 1.5-micron agent at a 500-micron separation distance, the power is highest in the bin centered about 2.0 MHz. Power is also higher in the bin corresponding to the second harmonic (4.0 MHz) compared to the other frequencies. The power in the 4.0-MHz bin noticeably decreases for initial radii greater than 1.5 microns.

With the 2-MHz driving pulse, a 3-micron agent is driven at twice its natural frequency, and this secondary resonance also results in higher power in the corresponding (2-MHz) frequency bin. Likewise, the bin containing the 1-MHz energy (0.75–1.25 MHz), which contains the $\frac{1}{2}$ harmonic response, has a high value of integrated power, second only to the frequency bin containing the driving frequency.

For a 1-micron initial radius, the power in the echo frequency bins corresponding to the fundamental (2 MHz) and second harmonic (4 MHz) of the driving frequency decrease substantially with a 10-micron separation compared to the 500-micron separation distance case. The middle harmonic observed near 3 MHz in Fig. 4(a) for a 10-micron separation distance and 1-micron initial radius contributes to the power of the bin centered at 3 MHz in Fig. 6(b). The power in this bin is comparable to those centered at 1 and 2 MHz. Thus, a small agent coupled to a larger one results in decreased echo power from the smaller agent and a decrease in its harmonic energy.

In Fig. 7, for the 0.5-MPa pressure amplitude, the decreased power from the coupling of a smaller agent to a larger one is also apparent. Notably, the 1.0-MHz bin is much lower for radii below 1.5 microns at the 10-micron separation distance compared with the 500-micron case. Also, as in Fig. 6, the total power is lower at the 10-micron separation distance. In Fig. 7(b) compared with Fig. 6(a), the 4.0-MHz bin decreases more sharply for radii from 1.5 to 2.0 microns. The corresponding 1.0- and 2.0-MHz bins rise more sharply. In this case, the bubbles are essentially uncoupled, and at the higher pressure amplitude, the second harmonic component shifts towards lower frequencies. In Fig. 6(b), the 4.0-MHz component is second highest at 1.5 microns, while it is the lowest in the corresponding plot at 0.5 MPa of Fig. 7(b). The driving at this level excites the higher-order nonlinearities in the coupling, thus diminishing the second harmonic response.

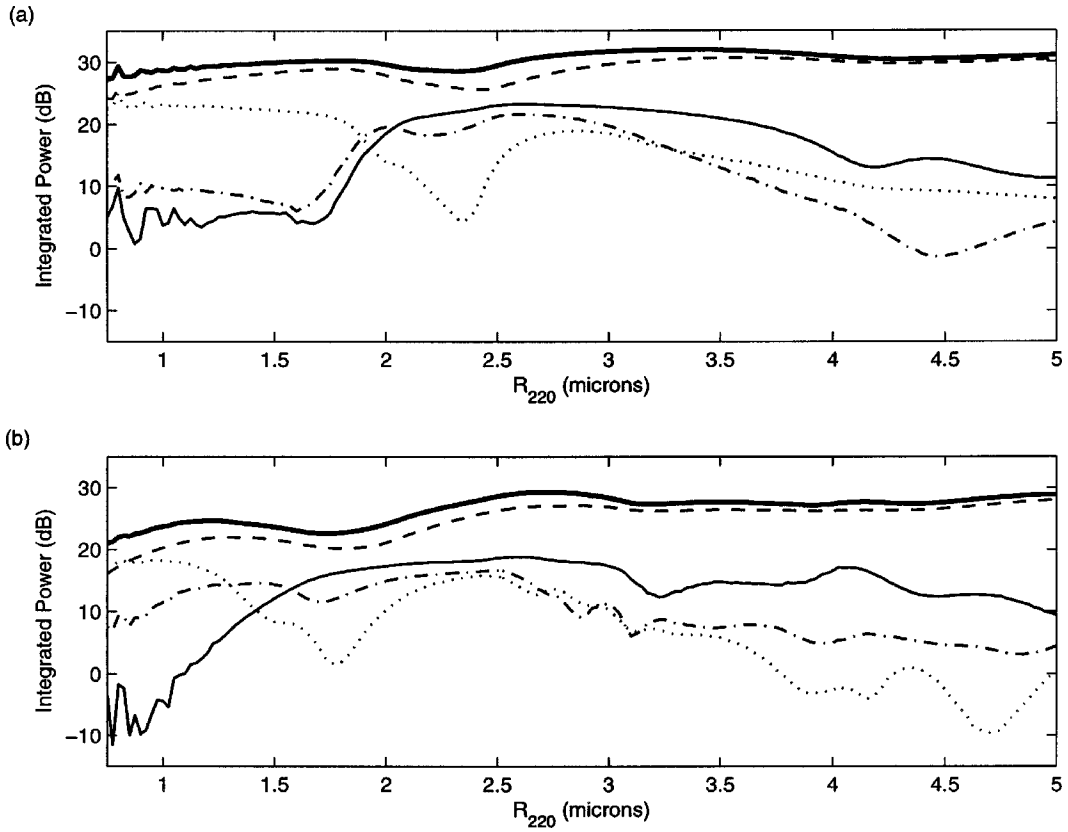


FIG. 7. (a) At a separation distance of 500 microns, integrated power is plotted over 500-kHz bins for a 2.2-micron agent coupled with second agent of initial radius from 1 to 5 microns for a 0.5-MPa acoustic pressure amplitude. The frequency bins are denoted as below 5 MHz (dark solid line), centered at 1.0 MHz between 0.75 and 1.25 MHz (solid line), centered at 2.0 MHz between 1.75 and 2.25 MHz (dashed line), centered at 3.0 MHz between 2.75 and 3.25 MHz (dash dot line), centered at 4.0 MHz between 3.75 and 4.25 MHz (dotted line). (b) At a separation distance of 10 microns, integrated power is plotted over 500-kHz bins for a 2.2-micron agent coupled with second agent of initial radius from 1 to 5 microns for a 0.5-MPa acoustic pressure amplitude. The frequency bins are denoted as below 5 MHz (dark solid line), centered at 1.0 MHz between 0.75 and 1.25 MHz (solid line), centered at 2.0 MHz between 1.75 and 2.25 MHz (dashed line), centered at 3.0 MHz between 2.75 and 3.25 MHz (dash dot line), centered at 4.0 MHz between 3.75 and 4.25 MHz (dotted line).

It should be noted that the 0.3-MPa pressure is frequently used for contrast agent imaging applications. At pressure amplitudes greater than this value, the nonlinear propagation effects of the tissue are likely to hamper detections schemes which assume tissue behaves like a linear scattering medium. Also, there is a much greater likelihood that the agent will be destroyed from the rupture of its shell.

B. Outstanding issues and applications

As mentioned earlier, previous work has assumed that contrast agents scatter independently such that any oscillation coupling is neglected. However, here it has been demonstrated that the radiated pressure may depend on the extent of coupling, which is a function of the separation distance for two agents forced above one mode and below the other. Under *in vivo* experimental and clinical conditions, coupling becomes an important issue. A typical 3-cc injection of 2.0-micron agents contains approximately $1-5\times10^{11}$ agents, depending on the concentration. For illustrative purposes, the agents are injected into a 5-L typical human blood volume and mixed homogeneously. Solving for the average volume that contains two agents yields interaction distances of 28.8 microns for 1×10^{11} agents and 16.8 microns for 5×10^{11} agents. The same calculation for agents at 1.0-micron radius ($8-40\times10^{11}$ agents) yields 14.4 and 8.4 microns, respectively. This order of magnitude calculation suggests that agents will be in close proximity on average such that the acoustic coupling between them should not be ignored. The manifestation of this effect in previous experiments has probably been attributed to increased attenuation.

Furthermore, contrast agent detection schemes have been developed based on the theoretical model of the nonlinear amplitude and phase response of a single agent.[1] Under moderate forcing conditions as a function of coupling distance, changes in the power in the second harmonic have been demonstrated in the simulations. The amplitude and phase of two or more agents may couple in a manner skewing the predictions made for single agents for certain forcing conditions and separation distances. New imaging or detection schemes might be developed which exploit these coupled responses. The nonlinear coupling between oscillations is a unique property that tissue does not exhibit and might be exploited for flow velocity detection.

For tissue perfusion studies and targeted drug delivery applications, the destruction of many agents in a localized area is desired.[2] Since agents in close proximity to each other may influence each other's radial responses, it likely that the acoustic parameters required for the destruction will be a nonlinear function of the agent's sizes, shell properties, and spacing. The pressure amplitude required to fragment a

single isolated agent may be significantly different than that required to fragment a group of agents. Likewise, any optimization of the acoustic destruction schemes will require an understanding of these coupled nonlinear effects.

The problem of bubble clouds and interactions continues to be a subject of continuing research for geophysical and industrial acoustic applications. It has received less attention and a limited treatment in the ultrasound contrast agent field. In these other fields the acoustic driving pressure amplitudes are usually sufficiently low that linearized versions of the equations can be used to obtain analytical results for direct comparison with experimental results.[12,21] Multiple scattering approaches have been used to examine the behavior of bubbles of the same size. However, with ultrasound contrast agents, a range of initial radii is expected. Likewise, harmonic imaging methods need higher-pressure amplitudes to drive agents nonlinearly, and the acoustically instigated destruction of the agents requires even higher pressure amplitudes. More theoretical work is needed to address these outstanding issues with respect to the experimental results. Rigorous compressibility corrections have been neglected in this work; however, their rigorous inclusion is needed for further study. A long-term goal is an understanding of the complicated nonlinear interactions of many agents for the ultrasound contrast agent applications of imaging and localized drug delivery.

VI. CONCLUSIONS

The equations for the coupled oscillations of two agents along their lines of center are generalized to include viscous fluid shells of arbitrary thickness. An energy-balanced system of equations may be obtained from this formulation that can simultaneously be solved along with the bubble dynamics equations. This provides a reasonable model for the coupled oscillations of lipid-shelled contrast agents within the constraints of incompressibility and a fixed separation distance. A numerical investigation examines the coupled oscillations of agents of differing radius with respect to their normal modes. The potential limitations of the assumptions of independent radial oscillations of multiple contrast agents in close proximity are investigated.

For a pair of agents with differing initial radius, the coupling at sufficiently small separation distances has a greater influence on the smaller agent. The inertial contribution of the coupling term may result in different collapse characteristics and a diminished radiated pressure. A decrease of 10 dB in the integrated scattered power from the smaller agent is predicted for the coupling of a 2.2-micron radius agent with a 1.0-micron radius agent at a 10-micron separation distance, compared to a 500-micron distance. The results have important implications for ultrasound contrast agent imaging and therapeutic methods where the nonlinearities of multiple agents cannot be avoided.

ACKNOWLEDGMENT

The authors acknowledge the support from the NIH CA 76062 and EB 00239.

[1] *Ultrasound Contrast Agents: Basic Principles and Clinical Applications*, edited by B. B. Goldberg, J. Raichelen, and F. Forsberg (Dunitz, London, 2001).
[2] E. Unger, T. P. McCreery, T. R. Sweitzer, S. DeKang, and G. L. Wu, ''*In vitro* studies of a new thrombus-specific ultrasound contrast agent,'' Am. J. Cardiol. **81**, 58G–61G (1998).
[3] E. Unger, T. McCreery, R. Sweitzer, V. Caldwell, and Y. Wu, ''Acoustically active lipospheres containing paclitaxel: A new therapeutic ultrasound contrast agent,'' Invest. Radiol. **12**, 886–892 (1998).
[4] J. S. Allen, D. J. May, and K. W. Ferrara, ''Dynamics of therapeutic contrast agents,'' Ultrasound Med. Biol. **28**, 805–816 (2002).
[5] C. C. Church, ''The effects of an elastic solid surface layer on the radial pulsations gas bubble,'' J. Acoust. Soc. Am. **97**, 1510–1521 (1995).
[6] L. Hoff, P. C. Sontum, and J. M. Hovem, ''Oscillations of polymeric microbubbles: Effect of the encapsulating shell,'' J. Acoust. Soc. Am. **107**, 2272–2280 (2000).
[7] S. Higenfeldt, D. Lohse, and M. Zomack, ''Sound scattering and localized heat deposition of pulse-driven microbubbles,'' J. Acoust. Soc. Am. **107**, 3530–3539 (2000).
[8] B. Anglesen, *Ultrasound Imaging: Waves, Signals and Signal Processing* (Emantec, Norway, 2001), Vol. II.
[9] V. F. K. Bjerknes, *Fields of Force* (Columbia, New York, 1906).
[10] X. Zheng and R. E. Apfel, ''Acoustic interaction forces between two fluid spheres in an acoustic field,'' J. Acoust. Soc. Am. **97**, 2218–2226 (1995).
[11] T. J. Leighton, *The Acoustic Bubble* (Academic, San Diego, 1994).
[12] C. E. Brennen, *Cavitation and Bubble Dynamics* (Oxford University Press, New York, 1995).
[13] Yi-Chun Wang and C. E. Brennen, ''Numerical computation of shock waves in a spherical cloud of cavitation bubbles,'' J. Fluids Eng. **121**, 872–880 (1999).
[14] N. A. Pelekasis and J. A. Tsamopoulous, ''Bjerknes forces between two bubble. II. Response to an oscillatory pressure field,'' J. Fluid Mech. **254**, 500–527 (1993).
[15] H. N. Oguz and A. Prosperetti, ''A generalization of the impulse and virial theorems with an application to bubble oscillations,'' J. Fluid Mech. **218**, 143–162 (1990).
[16] R. Mettin, I. Akhatov, U. Parlitz, C. D. Ohl, and W. Lauterborn, ''Bjerknes forces between small cavitation bubble in a strong acoustic field,'' Phys. Rev. E **56**, 2924–2927 (1997).
[17] E. A. Zabolotoskaya, ''Bjerknes force between bubbles,'' Sov. Phys. Acoust. **30**, 365–367 (1984).
[18] *High Intensity Ultrasonic Fields*, edited by L. D. Rozenberg (Plenum, New York, 1971).
[19] K. Morgan, J. Allen, P. Dayton, J. Chomas, A. Klibanov, and K. Ferrara, ''Experimental and theoretical evaluation of microbubble behavior: effect of transmitted phase and bubble size,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1494–1508 (2000).
[20] P. Dayton, J. Allen, and K. Ferrara, ''The magnitude of radiation force on ultrasound contrast agents,'' J. Acoust. Soc. Am. **112**, 2183–2192 (2002).
[21] C. Feuillade, ''Acoustically coupled gas bubbles in fluids: Time-domain phenomena,'' J. Acoust. Soc. Am. **109**, 2606–2615 (2001).
[22] V. V. Bredikim, A. Yu, and N. I. Vasilinenko, ''Auto-phasing of free volume oscillations of air cavities in water,'' J. Acoust. Soc. Am. **103**, 1775–1786 (1998).